East China - West China - New Zealand Tripartite Workshop

Computational Intelligence:

Methods, Systems and Applications for Ecological and Enviromental Modelling in China and New Zealand

Proceedings of the 1st Tripartite Workshop, Auckland, New Zealand

General Chair
Prof. Nikola Kasabov

Program Co-Chair
Dr. Raphael Hu

Local Organizer
Joyce D'Mello

Sponsors:

Ministry of Education, New Zealand
Ministry of Science And Innovation, New Zealand
Ministry of Foreign Affairs and Trade, New Zealand
Knowledge Engineering and Discovery Research Institue, New Zealand
Ministry of Science and Technology, China
Ministry of Education, China

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

KEDRI KNOWLEDGE ENGINEERING & DISCOVERY RESEARCH INSTITUTE

AUT UNIVERSITY

新疆大學

AUT University February 2011 KEDRI

Table of Contents

Welcome to the Tripartite Workshop 2011

The First Workshop of the Tripartite Programme between the Knowledge Engineering and Discovery Research Institute (KEDRI) at the Auckland University of Technology, New Zealand, Shanghai Jiao Tong University in Shanghai (East China) and Xinjiang University (West China) is a follow up on the signed tripartite agreement between the three partners Shanghai Jiao Tong University (SJTU), Shanghai, East China, Xinjiang University (XU), Urumqi, West China and Auckland University of Technology (AUT), New Zealand, and also on the discussions between the partners held in June 2010 during their meetings in China, and on the new initiatives of both the Chinese and the New Zealand Ministries of Education to support the tripartite joined research in the future.

The aims of the Workshop are:

- Presenting the progress to date of the research under the agreement.
- Introducing the Chinese partners to the methods and models developed at KEDRI for computational intelligence, including neural networks and quantum inspired evolutionary techniques; personalised modelling; multiple time series prediction.
- Introducing the Chinese partners to the ecological and environmental projects at the CORE Bio-protection, Lincoln and the research in this area in New Zealand.
- Discussing joint projects, including the establishment of a joint research centre in Urumqi.
- Discussing data and software to be shared and used for the joint research.
- Preparing a joint publication of the Proceedings of the Workshop and disseminating it to interested organisations in both NZ and China.
- Preparing a detailed working plan for the year 2011.

Auckland and Lincoln, 7-17 February 2011

Prof. Nikola Kasabov, Fellow IEEE, Fellow RSNZ
Director and Founder, Knowledge Engineering and Discovery Research Institute (KEDRI)
Chair of Knowledge Engineering, School of Computing and Mathematical Sciences,
Auckland University of Technology
Website: http://www.kedri.info

Workshop Presentation Slides

Intelligent Information Systems for Exploration, Understanding and Prediction of Ecological and Environmental Problems

Prof. Nikola Kasabov

SJTU – XU - AUT NZ
Tripartite Collaborative Programme

Workshop, February 2011, Auckland and Lincoln, NZ

Intelligent Information Systems for Exploration, Understanding and Prediction of Ecological and Environmental Problems

Prof. Nikola Kasabov, Fellow IEEE, Fellow RSNZ

Director, Knowledge Engineering and Discovery Research Institute, KEDRI
Auckland University of Technology

nkasabov@aut.ac.nz http://www.kedri.info

Teams Involved

- **Shanghai Jiao Tong University** (SJTU), lead by Prof. Jie Yang
- Ms. Xu Liming, SJTU

- **Xinjian University (XU)**, Urumqi, lead by Prof. Zhenghong Jia.
- Ms. Nan Zhang, XU

- **Auckland University of Technology (AUT),** Knowledge Engineering and Discovery Research Institute (KEDRI) lead by Prof. Nikola Kasabov
- Dr Raphael Hu – Deputy Coordinator
- Ms Joyce D'Mello – Administrative Manager

The KEDRI team on this Pogramme

- Prof. Nikola Kasabov – Co-ordinator
- Dr. Raphael Hu – Deputy co-ordinators - Personalised modelling systems
- Joyce D'Mello – Admin Manager
- Dr. Stefan Schliebs - Spiking neural networks (SNN)
- Dr Russel Pears - Methods for data mining.
- Dr. A.Mohemmed - Methods of smarm optimisation.
- M. Karaivanov (PhD) - Wind data modelling and wind energy decision support.
- H.Nuzly (PhD)- SNN and quantum inspired particle swarm optimisation.
- H.Widdiputra (PhD) - multiple time series prediction for environmental modelling.
- Ms. Linda Liang (PhD) – personalised modelling.
- Lei Song (PhD) – environmental modelling
- Shoba Tegginmath (PhD and lecturer) – Ontology systems
- Kshitij Dhoble (PhD) – SNN for spatio-temporal modelling.
- Associates:
  - A/Prof. Sue Worner - Lincoln University;
  - Dr Paul Pang - UNITEC

nkasabov@aut.ac.nz

KEDRI: The Knowledge Engineering and Discovery Research Institute at AUT (www.kedri.info)

- Established June 2002
- Funded by AUT, NERF (FRST), NZ industry
- External funds approx NZ$4 mln.
- 4 senior research fellows and post-docs
- 20 PhD and Masters students;
- 25 associated researchers
- Both fundamental and applied research (theory + practice)
- 220 refereed publications
- 5 PCT patents
- Multicultural environment (9 ethnic origins)
- Strong national and international collaboration

nkasabov@aut.ac.nz

NZ Universities:

•Lincoln University and the CoRE for Bioprotection
•University of Auckland
•University of Otago
•Massey University

NZ based research and commercial partners:

• PEBL
• FONTERRA
•TELECOM,
• CRI- GNS
• SCION/CRI-Forest Research Institute
• Middlemore Hospital
• Mighty River Power
• other

KEDRI's National and International Partners

Centre for Novel Methods of Computational Intelligence

Centre for Neurocomputing and Neuroinformatics

Centre for Bioinformatics

Centre for Data Mining and Decision Support Systems

Centre for Adaptive Pattern Recognition Systems

Centre for the Study of Creativity

International collaboration:

•**Bulgaria** - Tech. University Sofia, Plovdiv
• **China:** SJTU, XU, CASIA
•**Germany** – University of Kaiserslautern
• **Italy** - University of Trento
•**Japan** – NiCT- Tokyo; Kobe University; Kyushu Institute of Technology, RIKEN, Tamagawa University
• **Singapore** - Nanyang Technological University
•**Spain** – U. da Coruna
•**Switzerland** – ETH Zurich
•**UK** – University of Lancaster, University of Ulster, Imperial College
•**USA** - UC, Berkeley, Harvard University, National Cancer Institute, NIH
• **Vietnam** - NUV

nkasabov@aut.ac.nz

Aims of the Programme

(1) Collaborative research and development in information science, ecological and environmental modelling.
(2) Co-supervision of postgraduate students.
(3) Exchange of students and staff members involved in projects.
(4) Establishment of a joint laboratory (Research Center) for a long term research collaboration.
(5) Joined applications for research grants.
(6) Development of software systems for ecological and environmental event prognosis and their commercialisation in China and NZ.

nkasabov@aut.ac.nz www.kedri.infro

Research Projects of KEDRI for 2011 as part of the Tripartite Programme SJTU-AUT-XU

Project 1: Implement in software and hardware the KEDRI methods for *spatio-temporal pattern recognition* based on probabilistic evolving spiking neural networks (peSNN) and quantum inspired evolutionary computation and apply them for ecological data modelling in collaboration with the CORE Bioprotection Lincoln.

Project 2: Implement in software and hardware the KEDRI Integrated Optimisation Method for Personalised Modelling (IMPM) and apply it for ecological and environmental modelling in collaboration with the CORE Bioprotection Lincoln.

Project 3: Implement in software and hardware the KEDRI methods for multiple time series prediction and apply them for wind energy prediction and the prediction of extreme environmental events.

Project 4: In collaboration with XU and SJTU contribute to the establishment of a joined laboratory on *Computational Intelligent Systems* at XU.

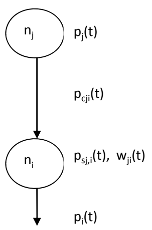
Expected results:

- Developed 3 software systems;
- Published 3 journal and 3 conference papers.
- Three PhD and 3 Masters students involved in the projects.

nkasabov@aut.ac.nz

Project 1: Implement in software and hardware the KEDRI methods for *spatio-temporal pattern recognition* based on probabilistic evolving spiking neural networks (peSNN) and quantum inspired evolutionary computation and apply them for ecological data modelling in collaboration with the CORE Bioprotection Lincoln.

Project co-leaders: N.Kasabov , S.Schliebs, Sue Worner
Participants: Ammar Mohhemed, Haza Nuzlu, Kshitij Doble, Nuttapod Nuntalid,



Probabilistic spiking neuron model, pSNM (*Kasabov, Neural Networks, Jan., 2010).*
The information in pSNM is represented as both connection weights and probabilistic parameters for spikes to occur and propagate. The neuron (n_i) receives input spikes from pre-synaptic neuron n_j ($j=1,2,\ldots,m$). The state of neuron n_i is described by the sum of the inputs received from all *m* synapses – the postsynaptic potential, PSPi(t). When PSPi(t) reaches a firing threshold $\vartheta i(t)$, neuron ni fires, i.e. emits a spike.

The PSPi(t) is now calculated using a new formula:

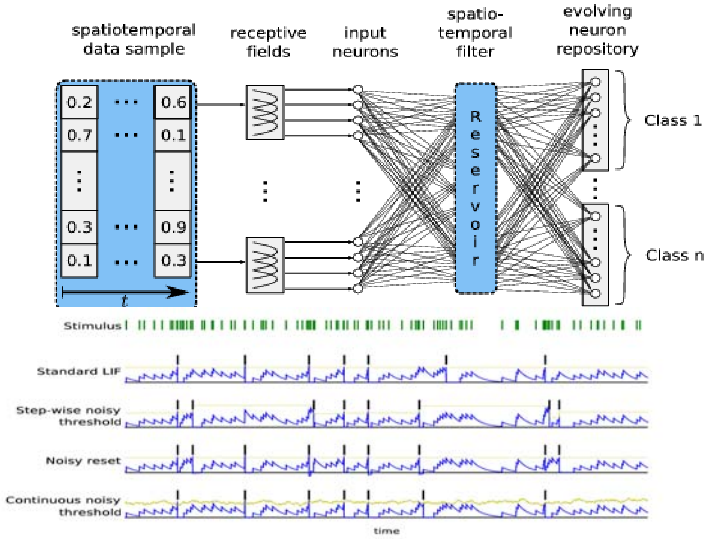
$$PSP_i(t) = \sum_{p=t_0,..,t} \sum_{j=1,..,m} e_j \, g(p_{cj,i}(t-p)) \, f(p_{sj,i}(t-p)) \, w_{j,i}(t) + \eta(t-t_0)$$

where: e_j is 1, if a spike has been emitted from neuron n_j, and 0 otherwise; $g(p_{cj,i}(t))$ is 1 with a probability $p_{cji}(t)$, and 0 otherwise; $f(p_{sj,i}(t))$ is 1 with a probability $p_{sj,i}(t)$, and 0 otherwise; t_0 is the time of the last spike emitted by n_i; $\eta(t-t_0)$ is an additional term representing decay in the PSP. As a special case, when all or some of the probability parameters are fixed to "1", the ipSNM will be simplified and will resemble some already known spiking neuron models, such as SRM.

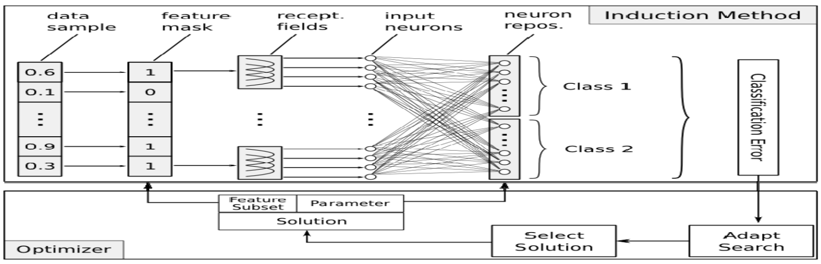
nkasabov@aut.ac.nz www.kedri.info

Software system of peSNN for STP recognition

S. Schliebs, N. Nuntalid, and N. Kasabov,Towards spatio-temporal pattern recognition using evolving spiking neural networks, Proc. ICONIP 2010, LNCS



Quantum inspired evolutionary optimisation of peSNN



1) The principle of quantum probability feature representation:
At any time a feature is both present and not present in a computational model, which is defined by the probability density amplitudes. When the model computes, the feature state is 'collapsed' in either 0 (not used) or 1 (used).

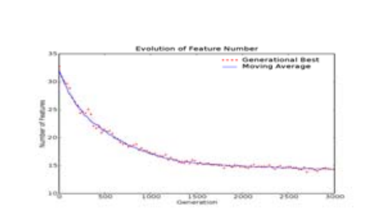
2) Quantum probability representation of the connections per the peSNN.

3) Quantum probability representation of the peSNN parameters.

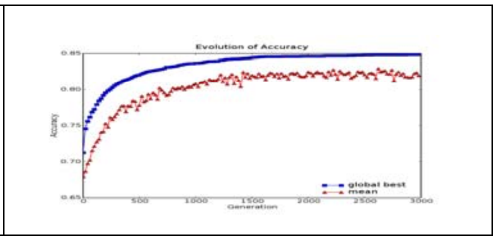
N.Kasabov, Integrative connectionist learning systems inspired by Nature: Current models, future trends and challenges, Natural Computation, Springer, 2009, 8:199-218.

An epSNN and QiEA for ecological modelling: Example on insect establishment prediction

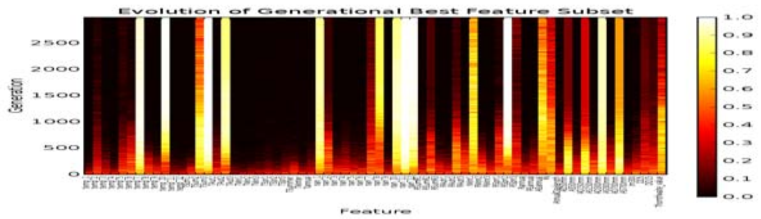
Participants: **S.Schliebs,** Haza Nuzlu, A/Prof. Sue Worner,, N.Kasabov



Evolution of classification accuracy on the climate data set after 3,000 generations.



Evolution of the features on the climate data set. The best accuracy model is obtained for 15 features .



Project 2: Implement in software and hardware the KEDRI Integrated Optimisation Method for Personalised Modelling (IPMP) and apply it for ecological and environmental modelling in collaboration with the CORE Bioprotection Lincoln.

Project co-leaders: N.Kasabov, Raphael Hu, Sue Worner

Participants: Linda Liang, Shoba Tegginmath

Recent publications: N. Kasabov and R.Hu, Integrated Optimisation Method for Personalised Modelling and Case Study Applications for Medical Decision Support, *J. Functional Informatics and Personalised Medicine*, in print, 2011

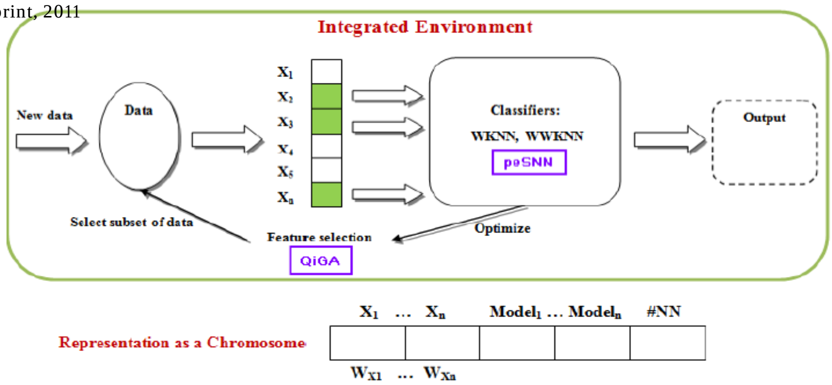
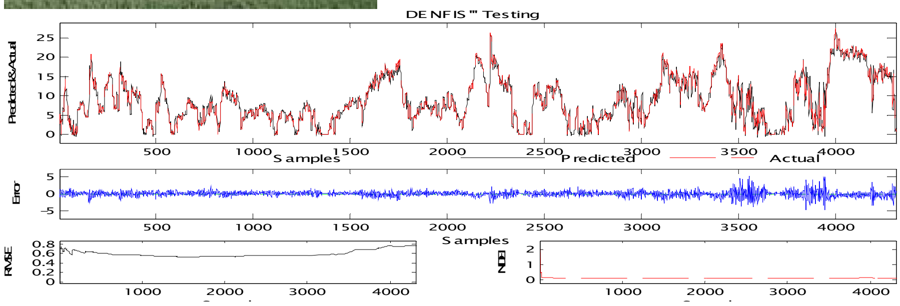


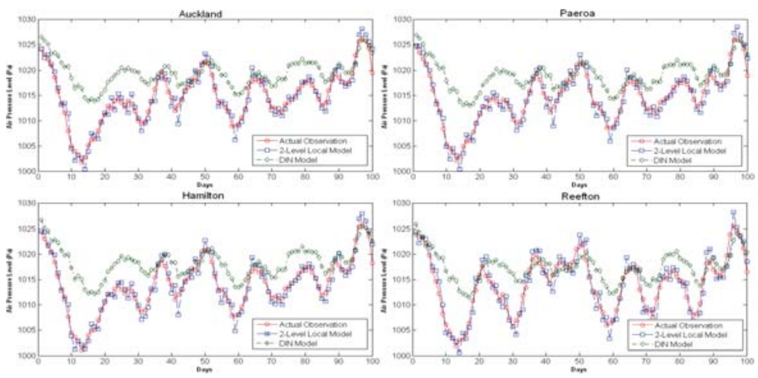
Fig.1: An overview of the novel personalised modelling system.

Project 3: Implement in software and hardware the KEDRI methods for multiple time series prediction and apply them for wind energy prediction and the prediction of extreme environmental events.

Project co-leaders: Russel Pears, N.Kasabov
Participants: Harya Widdiputra, M.Karaivanov



Intelligent information systems for the prediction of extreme environmental events



Example: results of 100 days (1st October 2009 to 31st December 2009) air pressure level prediction at four observation locations (Paeroa, Auckland, Hamilton and Reefton)

Computational Intelligence
and Modelling in Applied Ecology

Assoc. Prof. Susan P. Worner

Computational Intelligence and modelling in applied ecology

Susan P. Worner Takayoshi Ikeda, Dong Wang, Senait Senay, Hossein Khandan

Bio-Protection

Bioprotection science for New Zeala

Ecological applications

Bio-protection
Biosecurity
Food security

Invasive species

Prevention
Reducing impact of newly established species
Reducing impact of long established species

Invasive species: the effects

"Bioinvasion is fast becoming one of the world's most costly ecological problems". (Pimentel 2002)

Damage from alien plants, animals, and microbes in the United States, British Isles, Australia, South Africa, India, and Brazil account for more than $300 billion per year in direct damage and control costs. (Pimentel 2002)

- *Global travel, tourism and trade are increasing at an alarming rate*
- *Climate is changing*
- *Biofuels*
- *Food crisis*
- *All regions of the world face increasing threats from foreign invasive species*

Painted apple moth

Argentine ant

A NZ example: Painted apple moth

- *Pest of plantation forestry, horticultural and native forest trees*
- *Estimated impact of painted apple moth \$258M to \$500M over 20 years*
- *Costs of eradication 62M*

Many insects vector plant diseases

- *The glassy-winged sharpshooter is behind the spread of Pierce's disease among grapevines in California*
- *The insect infects vines with the bacterium Xylella fastidiosa when it feeds on the sap*

Ongoing costs of new invasive species

- ***Loss of sustainable production practices***
- ***New monitoring programs***
- ***New control strategies***
- ***Establishment of action thresholds***
- ***New pest control products and strategies***
- ***New biological control agents***
- ***Loss of trade opportunities***

Photograph by Raul Touzon

Parasitoid wasp laying eggs in an aphid

Biosecurity and quarantine agencies around the world are actively involved preventing new species invasion

The scale of the problem: a NZ example.

- *4000 - 5000 interceptions at the border of invertebrates per year in NZ*
- *Over 3000 potential global insect crop pest invaders*
- *The identification and prioritisation of new threats is difficult*

Chinese mitten crab

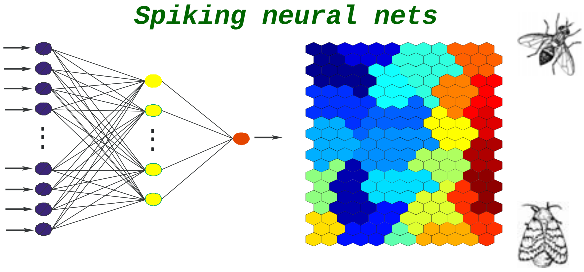
Asian tiger mosquito

Three components of pest risk assessment are addressed

1. *Which species?*
2. *Where?*
3. *Economic and environmental impact = how much damage?*

Computational intelligence (ecological informatics) methods

SOM's
MLP's
SVM's
Spiking neural nets

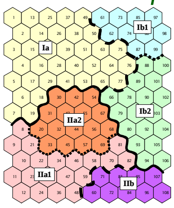


SELF ORGANISING MAPS

Can the pest profile in a region give useful information?

- The assemblage of species in a region integrates the complex factors influencing invasion

- Regions with similar pest grouping are assigned by a Self organising map (SOM) to nearby cells on the map



Database extract:

844 species (descriptors)

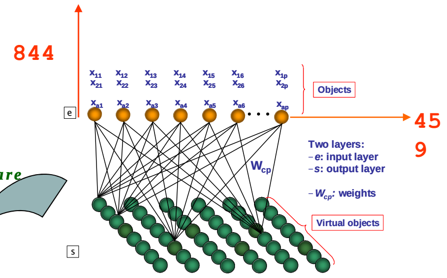
Species names and cells		Country names and codes								
		Armenia AM	India Bihar INbh	Italy IT	Jordan JO	Taiwan CNtw	-	-	-	Barbados BB
Abraxas grossulariata	ABRXGR	0	0	0	0	0	-	-	-	0
Acanthocoris scabrator	ACACS1	0	0	0	0	0	-	-	-	0
Acanthocoris scaber	ACACSC	0	0	0	0	0	-	-	-	0
Acanthocoris sordidus	ACACSO	0	0	0	0	1	-	-	-	0
Acaudaleyrodes rachipora	ACADCI	0	0	0	0	0	-	-	-	0
Achaea serva	ACAESE	0	0	0	0	0	-	-	-	0
Acanthiophilus helianthi	ACAIHE	0	1	1	1	0	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
Zeiraphera diniana	ZRPHDI	0	0	0	0	0	-	-	-	0
Zeiraphera rufimitrana	ZRPHRU	0	0	0	0	0	-	-	-	0
Zulia charon	ZULICH	0	0	0	0	0	-	-	-	0
Zulia colombiana	ZULICO	0	0	0	0	0	-	-	-	0
Zygaena filipendulae	ZYGNFI	0	0	1	0	0	-	-	-	0

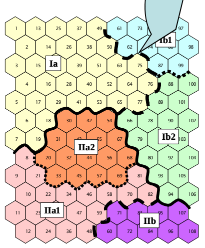
459 geographic areas (objects)

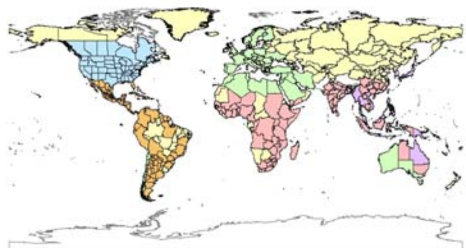
Crop Compendium – Global Module, 5th Edition © CAB International, Wallingford, UK, 2003

The Self Organising Feature Map (SOM)

SOM neural network orders 459 species assemblages projected onto map. Most similar pest groupings are assigned to the same cell



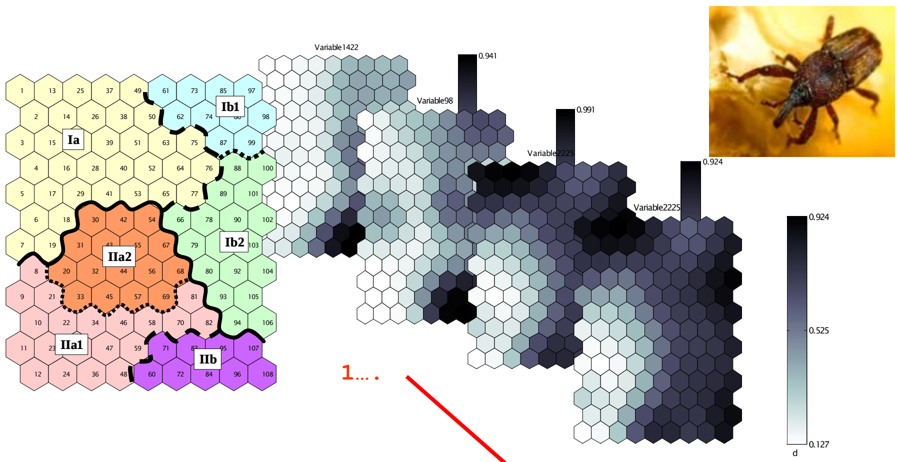




Project similar pest groupings onto world map

Some regions share a surprising number of species

Country	% Shared species	Similarity measure
Chile	52.5	0.87
Tasmania	50.0	0.92
South Australia	48.3	0.91
Victoria	48.3	0.91
Western Australia	45.0	0.89
Canary Islands	43.3	0.83
Tunisia	36.6	0.82
Malta	35.8	0.84
Azores	30.8	0.83
St Helena	28.3	0.80
Jordan	22.5	0.83



1….

844 Species

- Each species has its own map
- The strength of association of each species with regions and their species assemblages assigned to each cell is represented by the SOM weight

Worner SP & Gevrey M (2006) Modelling global insect pest species assemblages to determine risk of invasion. Journal of Applied Ecology 43, 858-867

Cydia pomonella	1	0.7654
Lampides boeticus	1	0.7619
Phthorimaea operculella	1	0.7585
Eriosoma lanigerum	1	0.7299
Delia platura	1	0.7261
Saissetia oleae	1	0.7129
Agrotis ipsilon	1	0.7123
Aphis craccivora	1	0.7079
Bemisia tabaci	1	0.7033
Rhopalosiphum maidis	1	0.7025
Acyrthosiphon pisum	1	0.6746
Saissetia coffeae	1	0.6652
Thrips tabaci	1	0.6612
Pseudococcus longispinus	1	0.6567
Locusta migratoria	1	0.6488
Aphis spiraecola	1	0.6479
Agrotis segetum	0	0.6432
Aspidiotus nerii	1	0.6425
Ceratitis capitata	0	0.6266
Hyperomyzus lactucae	1	0.6263
Phyllocnistis citrella	0	0.6228
Rhopalosiphum padi	1	0.6168
Spodoptera exigua	0	0.6165
Pieris brassicae	0	0.6009
Ephestia elutella	1	0.5703

Species list for New Zealand

Weights

1

Established species

0

Not Established species

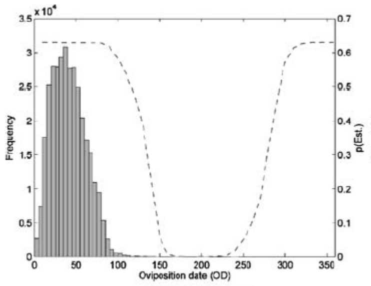
Pest risk assessment

1. Which species?

2. Where?

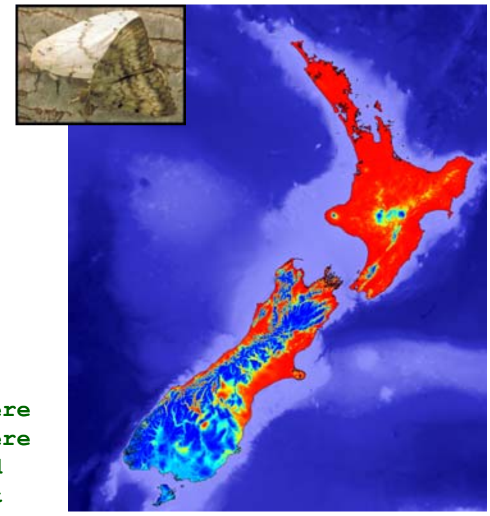
3. Economic and environmental impact

Likelihood of establishment of gypsy moth using a detailed mechanistic model

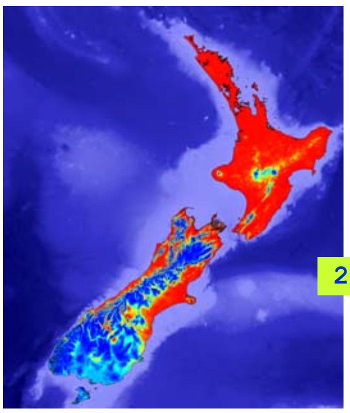


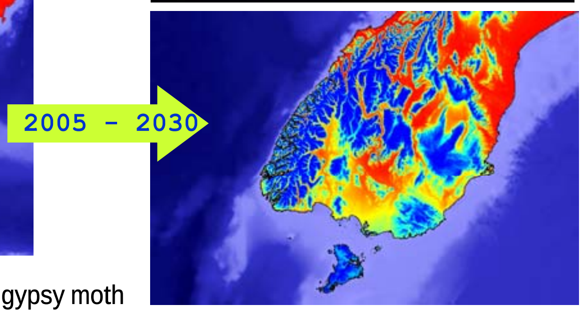
a) New Zealand

Weather and elevation data were used as inputs to predict where sustainable populations could potentially exist and predict the timing of hatch and oviposition in different regions.



Pitt J, Regniere J, & Worner SP (2006) Risk Assessment of gypsy moth, *Lymantria dispar* (L.) in New Zealand based on phenology modelling. *International Journal of Biometeorology* 51: 295-305



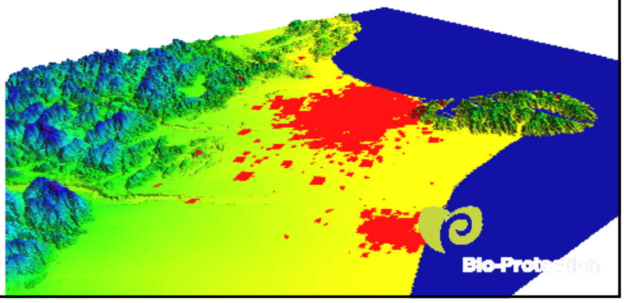


Likelihood of establishment of gypsy moth in response to climate change

Pitt J, Regniere J, & Worner SP (2006) Risk Assessment of gypsy moth, *Lymantria dispar* (L.) in New Zealand based on phenology modelling. *International Journal of Biometeorology* 51: 295-305

Predicting the potential distribution of little known invasive species

- *Decision makers need to make rapid decisions when little is known about a species biology and ecology*
- *Some measure of habitat suitability is required*
- *Many different modelling methods are available*



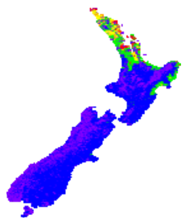
Species distribution, Niche modelling Habitat suitability modelling

For input, the models use a set of point localities where the species is known to occur and a set of geographic layers representing the environmental parameters that might limit the species' capabilities to survive.

+

Temperature
Rainfall
Evaporation
Soil moisture
Landcover

=



Species occurrence ***Environmental data*** ***Habitat suitability***

Computer programs may or may not include statistical modelling approaches

BIOCLIM/ANUCLIM,
BioMOD
CLIMATE
CLIMATE ENVELOPE RANGE
CLIMEX
DOMAIN
ENFA
GARP
GRASP
HABITAT
MaxEnt
NAPPFAST
STASH
OPENMODELLER

Modified from AWPRM 2007 Report October 15 2007.

Conventional statistical models, machine learning, statistical learning or supervised learning models

GLIM/GAM
Logistic regression
Naïve Bayes
Multivariate adaptive regression splines
Classification and Regression tree analysis
Artificial Neural Networks (Multi-layer perceptrons)
Support Vector Machines
Ensemble models

Modern statistical learning methods:

- ***are easily validated against independent data***
- ***error and uncertainty can be estimated***

Gevrey M, & Worner SP (2006) Prediction of Global Distribution of Insect Pest Species in Relation to Climate by Using an Ecological Informatics Method. *Journal of Economic Entomology* 99: 979 – 986.

Model comparison to predict species presence/absence:

Models

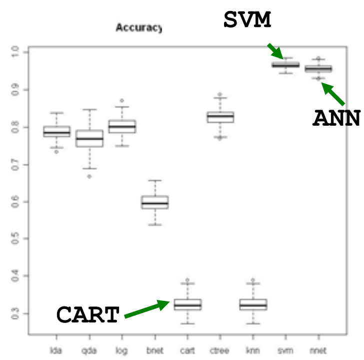
Logistic discriminant analysis
Quadratic discriminant analysis
Logistic regression
Naive Bayes
Classification and regression trees
Conditional trees
K-nearest neighbours
Support vector machines
Artificial neural networks

Performance measures

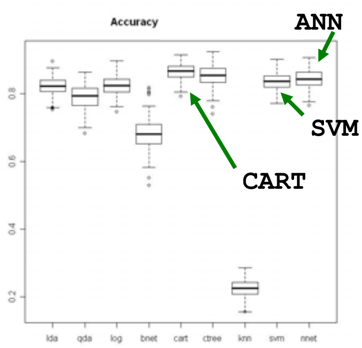
Accuracy
Precision
Recall
F-score
Kappa
Specificity
True skill statistic
AUC ROC
Uncertainty
0.632+ error

***Data:* Worldclim** *data and* **Bioclim** *abiotic variables were used to fit/develop models to determine habitat suitability for a number of species*

Chrysomphalus dictyospermi



Ceratitis capitata



1000 bootstrap samples

Validation on "out of bag" data

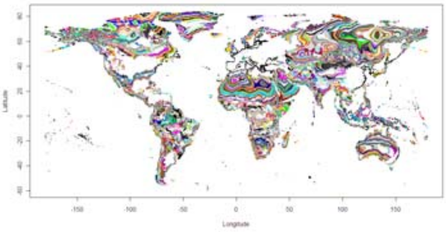
Variable model performance across species

Data characteristics are important especially absence data

1. *Use an ensemble of one class SVM to determine areas of environmental dissimilarity (prob. presence=0)*

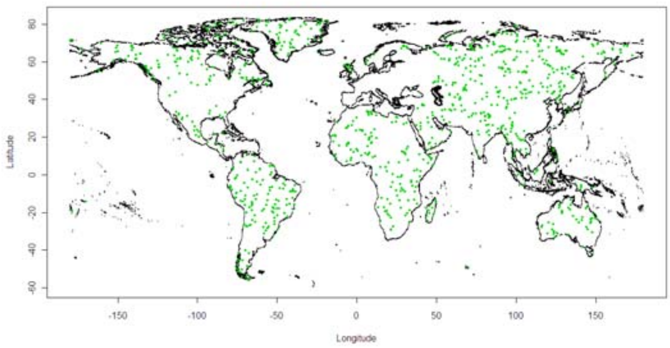
2. *High resolution geo-referenced data results in many thousands , ~400,000, locations*

3. *Cluster data (k-means) by choosing the number of clusters to balance presence locations (912)*



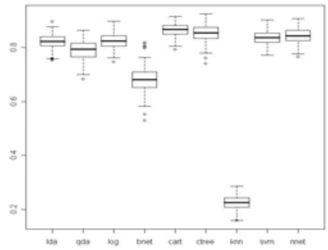
Clusters of dissimilarity (912) indicated by different colours, of environmental conditions that are most dissimilar to areas where Chinese mitten crab is likely to be present (white areas).

Choose centroids of those clusters to represent absence locations

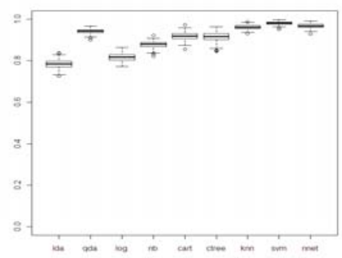


Absences derived from the centroids of the 912 clusters that balance the 912 recorded presences for Chinese mitten crab.

Improved Model Performance - Accuracy



Before



After

Predicting global distribution of mediterranean fruit fly

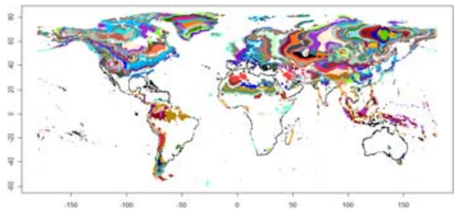


Figure 3.32:Clusters of locations (138) with similar environmental conditions most likely to be unsuitable for *C. capitata* establishment and therefore represent absent points Locations in white represent potential presence points.

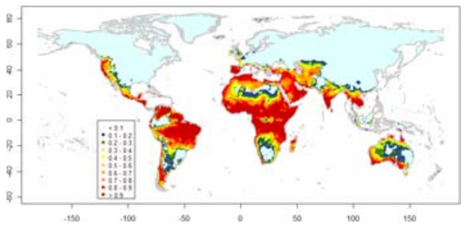


Figure 3.36: Global environmental suitability of *C. capitata* using a support vector machine
The legend represents environmental suitability with warmer colours representing higher environmental suitability.

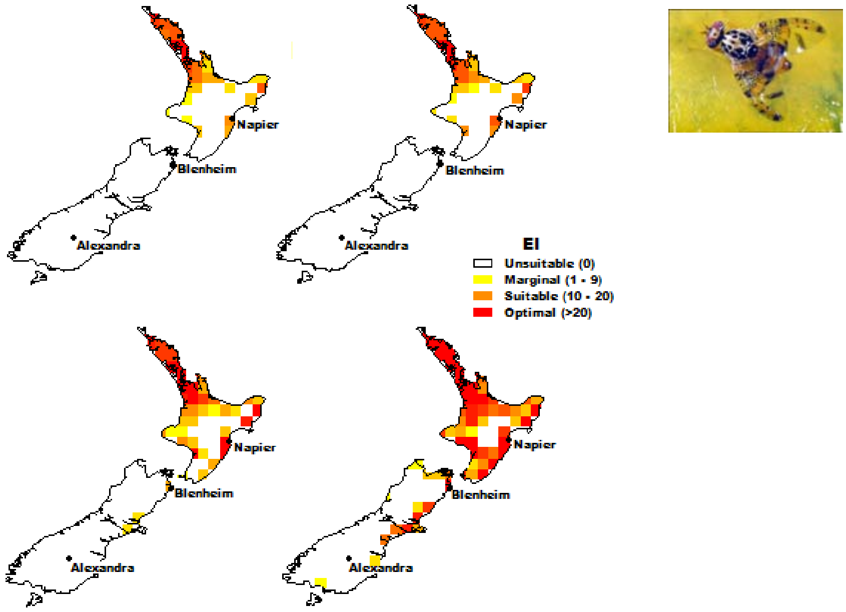
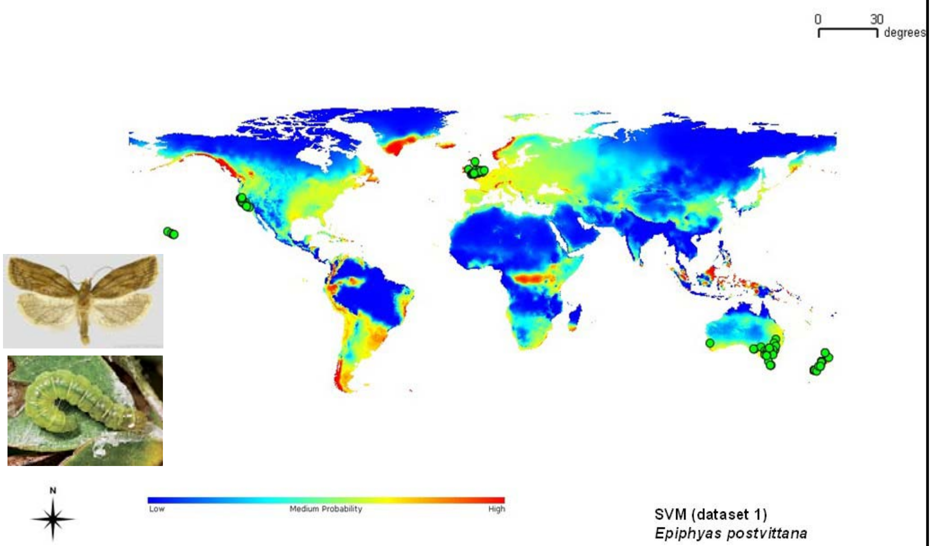
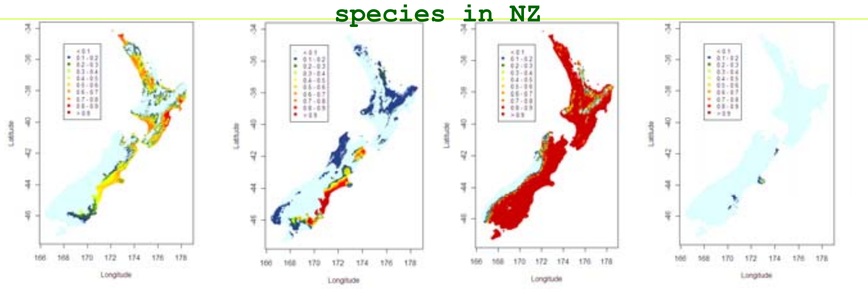


Figure 2.12: Potential distribution of *C. capitata* in New Zealand current climate (1961-1990 average), best case 2040 climate change predictions, average 2040 climate change predictions and worst case 2040 climate change scenarios for New Zealand.

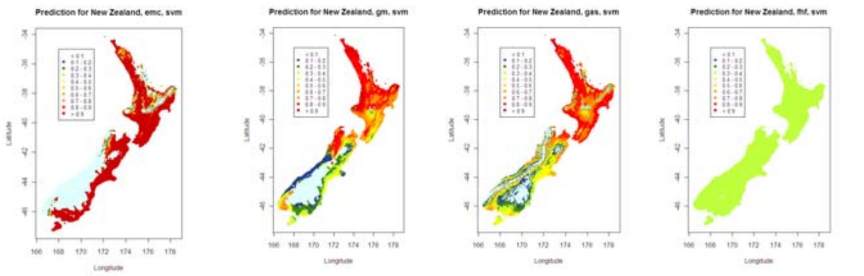
Predicted potential global distribution of *light brown apple moth* using an SVM



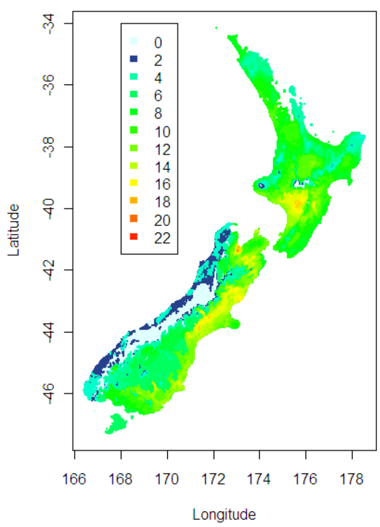
Probability of establishment of freshwater invasive species in NZ



Asian clam | Asian Calanoid copepod | Gammarus trigrinus | Chinese mitten crab



Estuarine mud crab | Golden mussel | Golden apple snail | Fishhook flea



Combine maps for 22 high risk freshwater species to identify hotspots

Bio-Protection

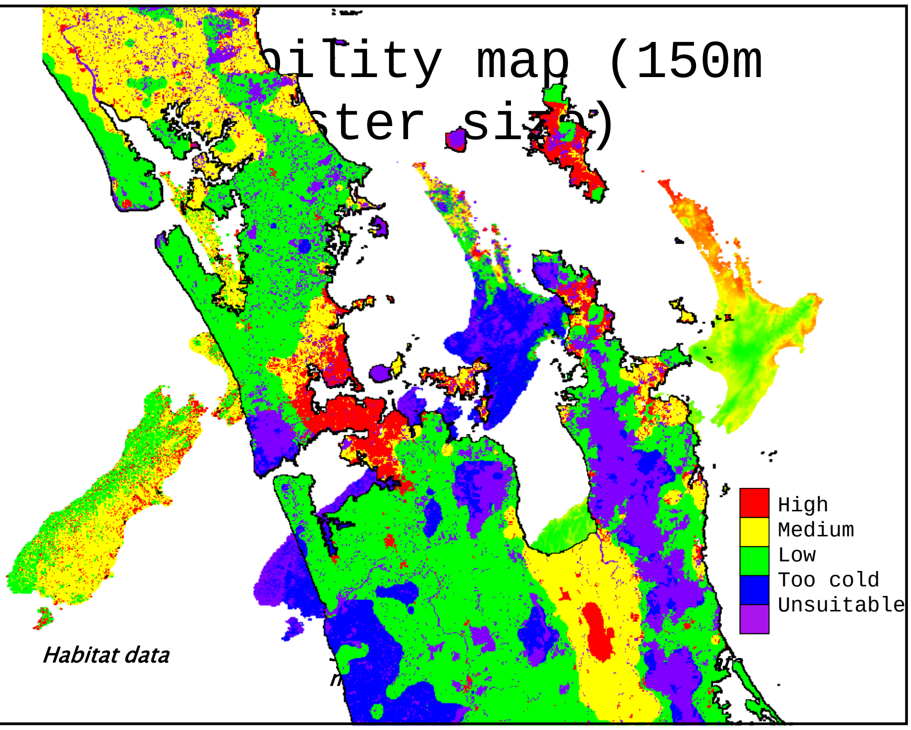
BIOSECURITY NEW ZEALAND

Pest risk assessment

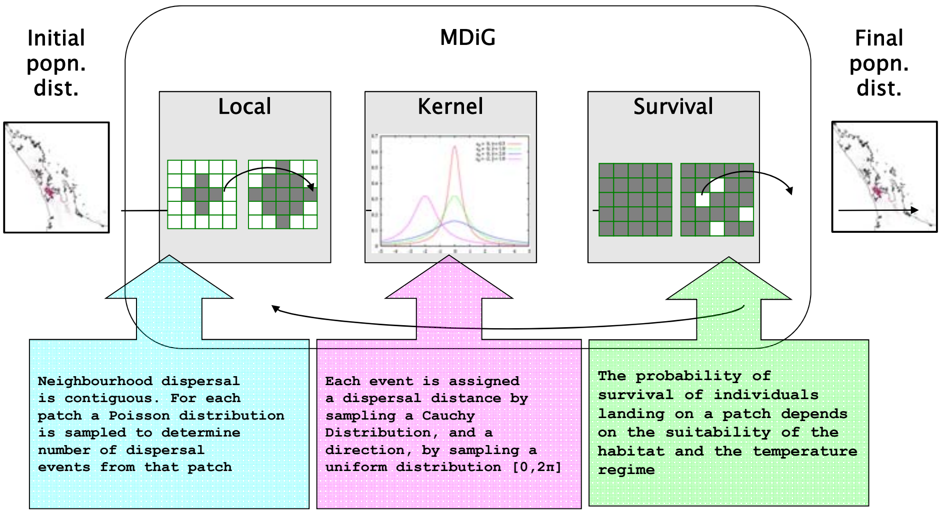
1. Pest categorisation =which species

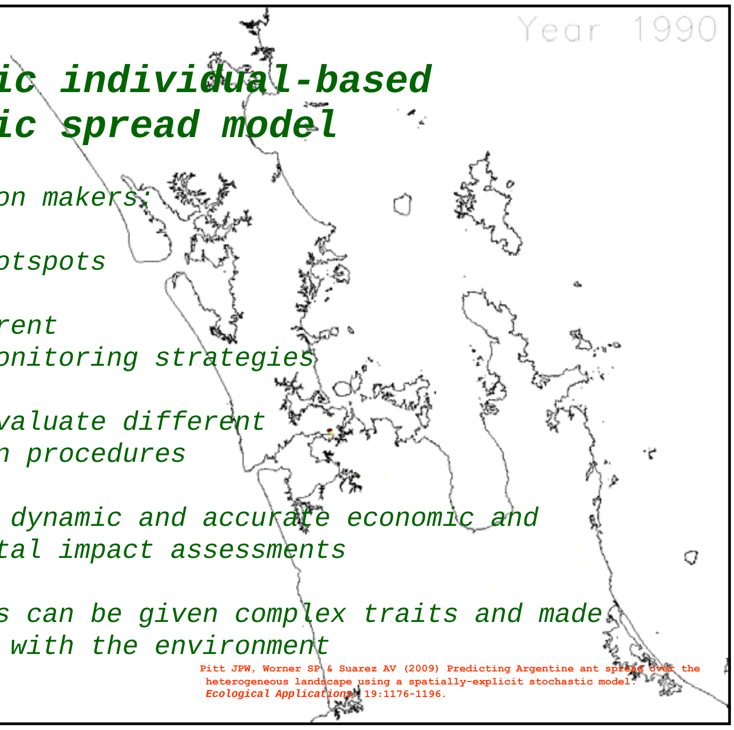
2. Risk of establishment = where

3. Economic and environmental impact



Individual based dispersal model





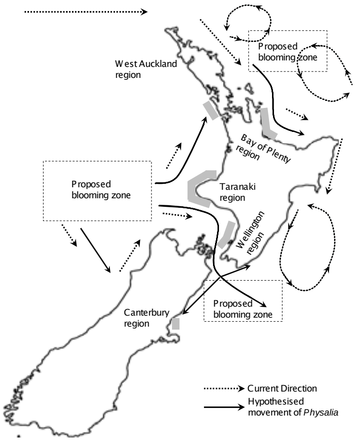
Stochastic individual-based stochastic spread model

Helps decision makers:

- *identify hotspots*
- *test different sampling/monitoring strategies*
- *test and evaluate different eradication procedures*
- *make more dynamic and accurate economic and environmental impact assessments*
- *individuals can be given complex traits and made to interact with the environment*

Pitt JPW, Worner SP & Suarez AV (2009) Predicting Argentine ant spread over the heterogeneous landscape using a spatially-explicit stochastic model. *Ecological Applications*. 19:1176-1196.

Predicting stinging jellyfish arrival on swimming beaches



Quantum-inspired Evolutionary Algorithm in conjunction with an Naive Bayes classifier to identify and clarify oceanographic features and time frames that influence *Physalia* dispersal around the New Zealand shoreline.

Bio-Protection

Models for Classical pest control and biological control

- An artificial neural network (ANN) model to forecast *Rhopalosiphum padi* abundance based historical trap catch records and weather data
- ANN model is compared with multiple regression
- Jackknifing is used to validate the models
- A two stage ensemble is used to improve prediction

Stage 2 Ensemble

For the new year train networks
y_{p1} , y_{p2} , y_{p3}, … … … y_{pn}

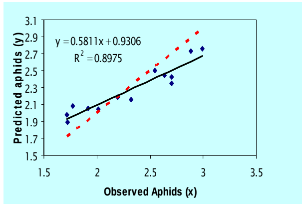
Lowest minimum average error | Highest R^2

y_{p1} y_{p2} y_{p3} y_{p4}

$x_{p1}=(y_{p1}-b)/a$ $x_{p1}=(y_{p1}-b)/a$ $x_{p1}=(y_{p1}-b)/a$ $x_{p1}=(y_{p1}-b)/a$

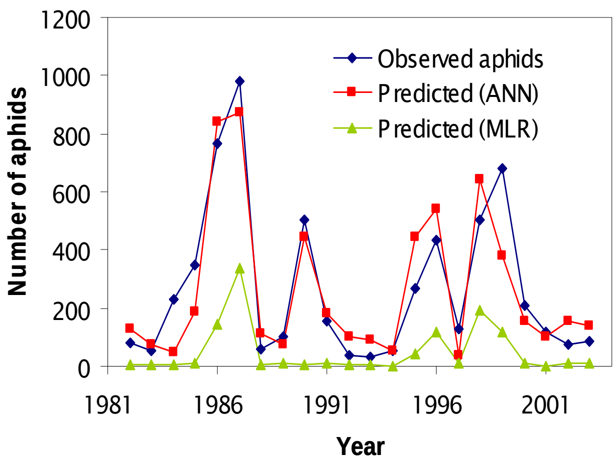
$$X_p = \frac{x_{1p}+x_{2p}+x_{3p}+x_{4p}}{4}$$

Linear ensemble



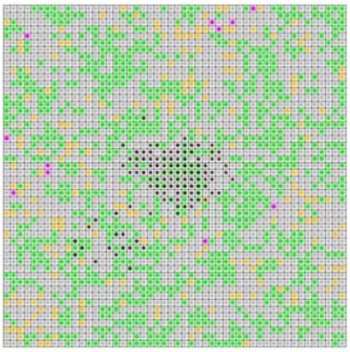
Application in pest control

Aphid flights predicted over 22 years using an ensemble of MLP's

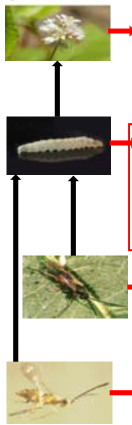


Leave one out validation - The data for each year is left out of the data set, a model is fitted to the remaining data and then used to make a prediction for the year not included in the dataset

Agent based model: Bio-Control



2D
Grid

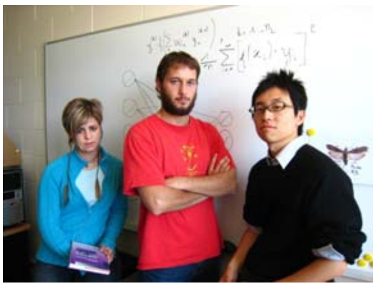


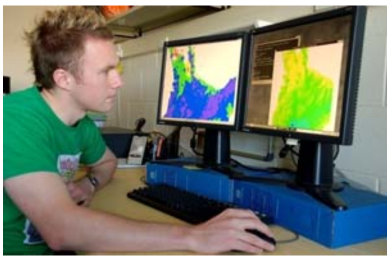
Host Plant Agent alive
Host Plant Agent dead

Pest Agent Alive,
Dying Pest Agent,
Attribute Energy &
Age

Parasitoid Agent
Attribute Energy
& Age

Predator Agent,
Attribute Energy
& Age






Thankyou
http://www.bioprotection.org.nz
Bio-Protection

Key Technology Research on Feature Analysis and Information Extraction of Remote Sensing Assessment Ecosystem

Prof. Jie Yang

Shanghai Jiaotong University

Key Technology Research on Feature Analysis and Information Extraction of Remote Sensing Assessment Ecosystem

Reporter: Prof. Jie Yang
Shanghai Jiaotong University

Institute of Image Processing & Pattern Recog

Contents

1 Introduction

2 System Framework

3 Change Detection

4 Image Classification

5 Information Extraction

6 Work Conclusion

1. Introduction

1.1 Basic introduction

In process of our research, images of Landsat Thematic Mapper (Landsat TM) are collected from Changji city, Xinjiang Province, China, and exist in the type of multi-band image(from band-1 to band-7).

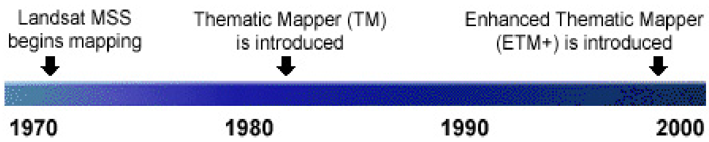
The feature of multi-band Landsat TM image from Changji city, Xinjiang province, China, must be known before starting the research. It concludes Landsat TM image, Multi-band remote sensing image, and the terrain of Changji city.

Landsat TM image	Landsat Thematic Mapper (TM) is a multispectral scanning radiometer that was carried on board Landsats 4/5 and 7(Landset TM7,ETM).
Multi-band RS image	The remote sensing image (Band 1-7).
Regionalism in Changji, China	The regional features in Changji, Xinjiang are very valuable for remote sensing research. We can predict the ecological change.

1.2 Landsat TM

- For 30 years, Landsat TM goes through three stages:
- In the early 1970s, with a series of 3 MSS satellites that mapped using the 3 visible channels along with a near-infrared band.
- In the early 1980s, Landsat satellites began offering what is known as the Thematic Mapper, which added two more infrared bands and a thermal long-wave infrared band, and doubled the resolution capabilities of the multispectral bands.

- In the 3rd stage, Landsat 7 successfully debuted the Enhanced Thematic Mapper (ETM+) platform in 1999. The ETM+ platform provides a continuation of the medium resolution imagery of a quality and affordability that made TM popular. It also added a higher resolution panchromatic band for aiding in interpretation.



Band information

- **Band 1(0.45-0.52um):** provides increased penetration of water bodies and also capable of differentiating soil and rock surfaces from vegetation and for detecting cultural features.
- **Band 2(0.52-0.60um):** sensitive to water turbidity differences; it highlighted the turbid water in the Barkley Lake and has separated vegetation (forest, croplands with standing crops) from soil.
- **Band 3 (0.63-0.69um):** senses in a strong chlorophyll absorption region and strong reflectance region for most soils. It has discriminated vegetation and soil. This band has highlighted barren lands, urban areas, street pattern in the urban area and highways. It has also separated croplands with standing crops from bare croplands with stubble.

- **Band 4(0.76-0.90um):** Well distinguish vegetation varieties and conditions; delineate water bodies, distinguished between dry and moist soils, and also separated croplands from bare croplands; is useful for crop identification and emphasizes soil-crop and land-water contrast.
- **Band 5(1.55-1.75um):** sensitive to the turgidity or amount of water in plants; can separate forest lands, croplands, water body distinctly. also can separate water body (dark tone) from barren lands, croplands, and grass lands (lighter tone).
- **Band 7(2.08-2.35um):** has separated land and water sharply; has strong water absorption region and strong reflectance region for soil and rock. Urban area, croplands, highways, bare croplands have appeared as bright tone and water body, forest have appeared as dark tone

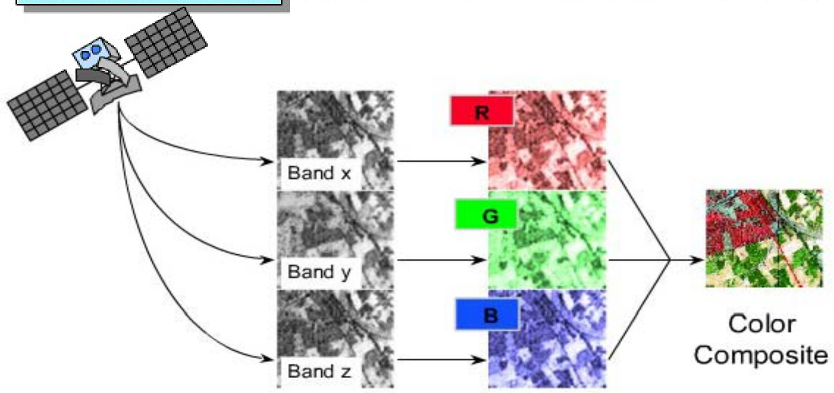
Spectral Resolution

Band	TM	ETM+	TM	ETM+
1 (Blue)	30 m	30 m	0.45-0.52 µm	0.45-0.52 µm
2 (Green)	30 m	30 m	0.52-0.60 µm	0.53-0.61 µm
3 (Red)	30 m	30 m	0.63-0.69 µm	0.63-0.69 µm
4 (Near IR)	30 m	30 m	0.76-0.90 µm	0.78-0.90 µm
5 (Middle IR)	30 m	30 m	1.55-1.75 µm	1.55-1.75 µm
6 (Thermal IR)*	120 m	60 m	10.4-12.5 µm	10.4-12.5 µm
7 (Middle IR)	30 m	30 m	2.08-2.35µm	2.09-2.35 µm
8 (Panchromatic)**		15 m		0.52-0.90 µm

ETM+ Band 6 (Thermal IR) includes both high and low gain setting
ETM+ Band 8 (Panchromatic) - most visible & near-IR data in sing

1.3 Multi-band remote sensing image

MULTI-BAND DISPLAY Color results from the ***compositing*** of 2 or 3 bands



NOTE : The bands can be individually *stretched*

MULTI-BAND PRODUCTS

Standard RGB color composite

NIR band	⇨	RED
RED band	⇨	GREEN
GREEN band	⇨	BLUE

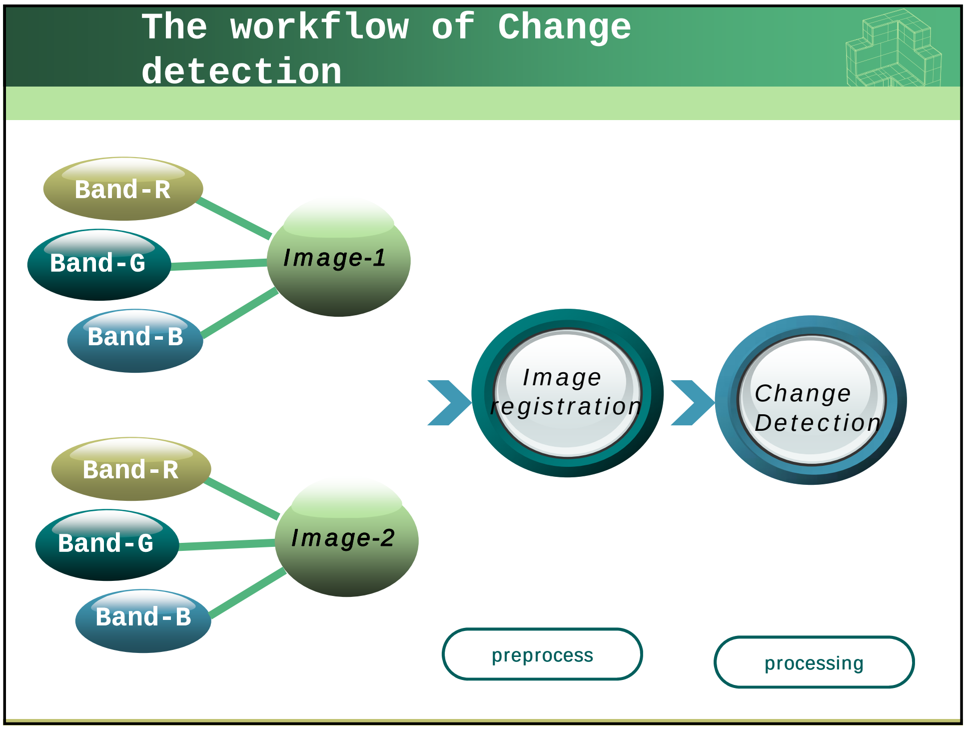
Example: SPOT XS3, XS2, XS1

2. System Framework

2.1.Framework workflow

- In our framework, two schemas are proposed to solve the problems:

Change Detectio n	Multi-band remote sensing image: Firstly, synthesize different band images and generate a proprocessed remote sensing images; Then, apply our method to process them.
Information Extraction	We can extract the information including the vegetation, water body, bare soil, and residential area at each year. Then we can detect the change region.


The workflow of Change detection
Band-R
Band-G
Band-B
Image-1
Band-R
Band-G
Band-B
Image-2
Image registration
Change Detection
preprocess
processing

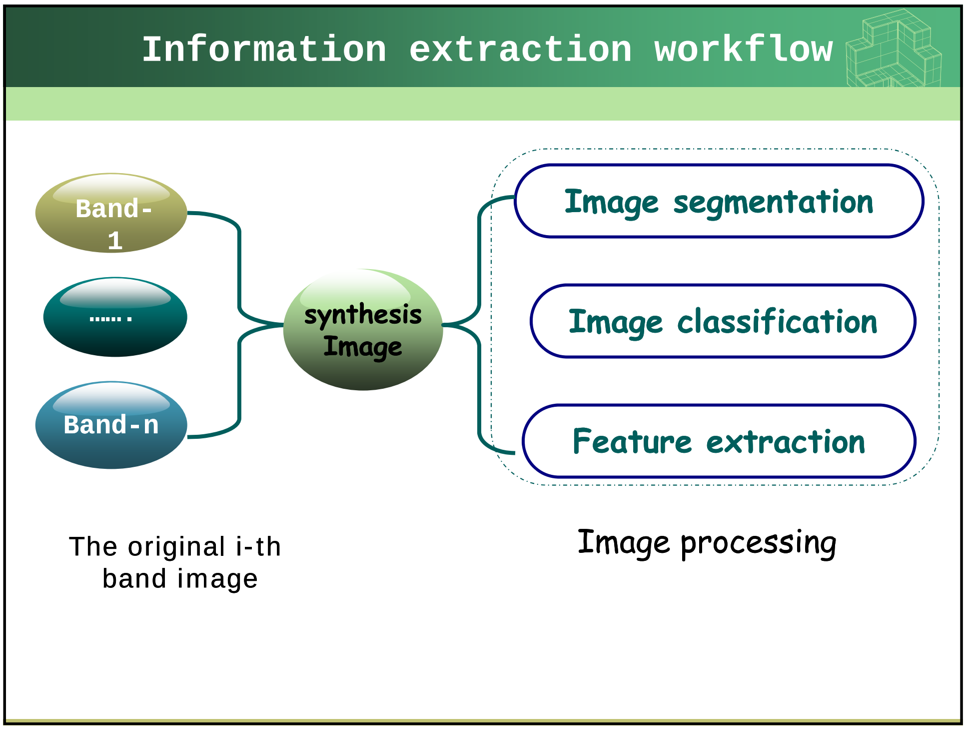
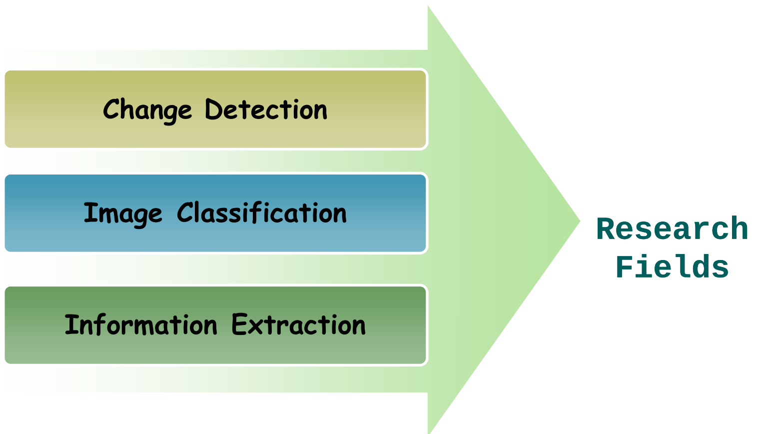

Information extraction workflow
Band-1
……
Band-n
synthesis Image
Image segmentation
Image classification
Feature extraction
The original i-th band image
Image processing

Image processing modules



3. Change Detection

3.1 Introduction

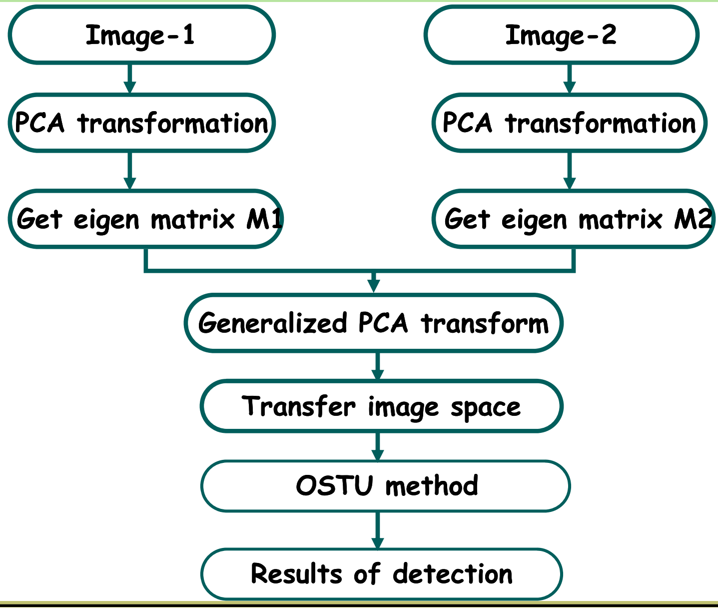
- In the image change detection, it must be registered and transformed before processing:

Image Registration & Transformati	For two images from different years, it must be registered and transformed before processing. In our development, SIFT and transformation equations are used to solve this problem.
Change detection	Firstly, two image must be synthesized from multi-band. Then, the improve PCA method is applied to detect the changed regions.

3.2 Scale Invariant Feature Transform

- Scale Invariant Feature Transform (SIFT): Proposed by David Lowe, its idea is to transform image data into scale-invariant coordinates relative to local features by matching by stable, robust and distinctive local features.
- Scale-space extrema detection
- Keypoint localization
- Orientation assignment
- Keypoint descriptor

3.3 Change detection



4. Image Classification

4.1 The methods of image classification

- Supervised
  - Maximum Likelihood
  - SVM
- Unsupervised
  - ISOData
  - K-Mean
- Semi-supervised
  - Graph-based

- *Supervised*: Labeled instances are often difficult, expensive, or time consuming to obtain, and require experienced human annotators.

- *Unsupervised:* Uses no prior knowledge to improve results, and often, needs to adjust many parameters; less accurate.

- *Semi-supervised* learning uses large amount of unlabeled data, together with the labeled data to build better classifiers.

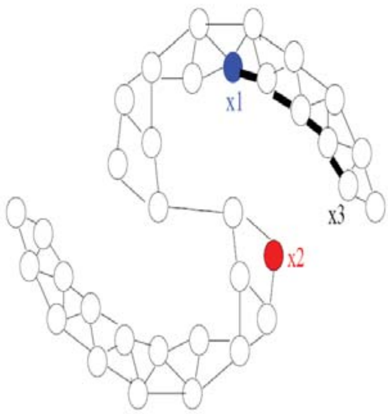
4.3 Support Vector Machine (SVM)

Principle:

SVM is a classical method and robust for data classification. The principle is well-known for many researchers and here is shown the basic idea using a figure:

4.5 Graph-based semi-supervised Classification

Graph Construction



Construct Graph $G = (V,E)$

vertices: $V = \{x_1, x_2, \ldots, x_n\}$

Edges: $E : \mathbf{W} = \{w_{ij} \geq 0\}$

Graph types:

1)Weight: $w_{ij} = \exp(\| x_i - x_j \|^2 / 2\sigma^2)$

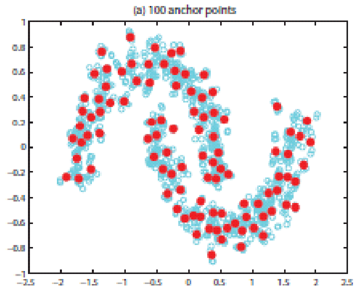
2)Knn: $w_{ij} = 1, \ if \ x_j \in \{Knn \, of \ x_i\}$

3)$\mathcal{E}$nn: $w_{ij} = 1, \ if \ \| x_j - x_i \| < \varepsilon$

Algorithm 2 - AnchorGraph

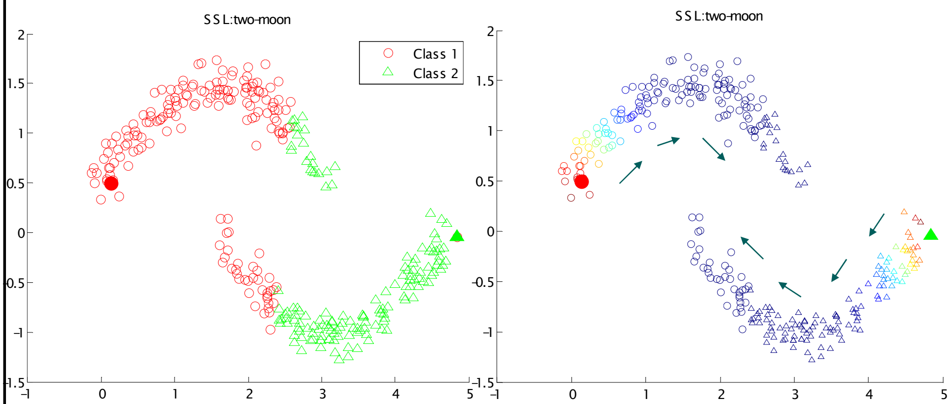
- General Principle: using a subset of data, i.e. **anchor points**, to design the graph and learn the function, then making prediction on full dataset by weighted combination of anchors!

$$f(\mathbf{x}_i) = \sum_{k=1}^{m} Z_{ik} f(\mathbf{u}_k)$$



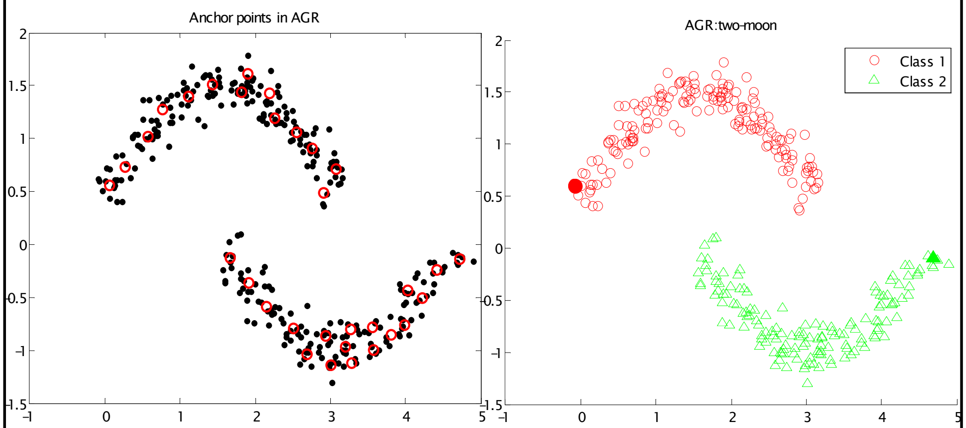
Experimental results

Iteration method



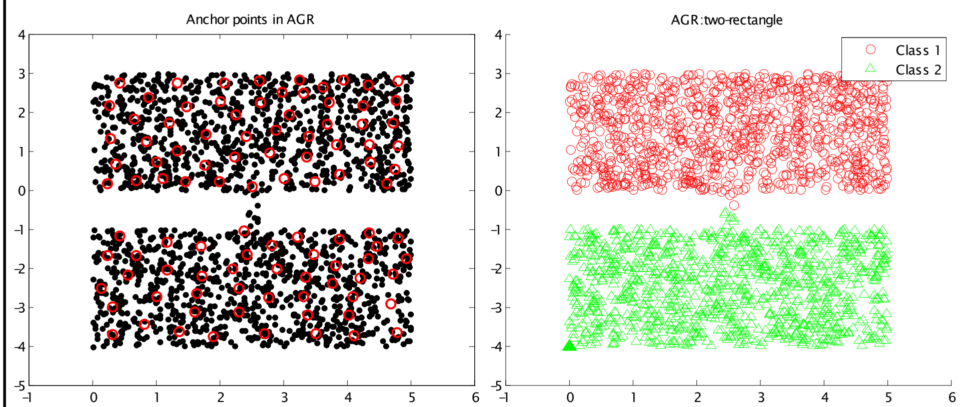
(a) Label prediction (b) energy distribution

loop=1

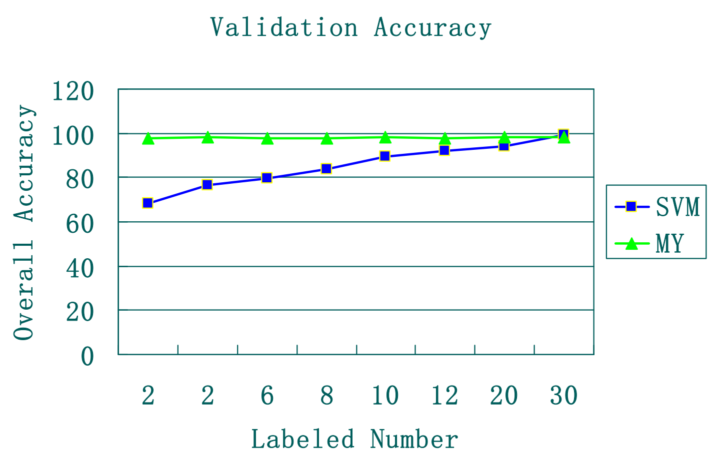


(a) Anchor points by K-mean (b) label predication

Anchor Graph Regularization



(a) Anchor points by K-mean (b) label predication



Comparison

	ML	SVM	OGR	AGR	MY
Min mem.	linear	linear	46GB	linear	linear
Time	18.1s	0.27s*	>5h	>1000s	1.66s

NOTES:

•Computer configuration: 2x2.66GHZ, 2GB RAM; MATLAB 2010a

•Evaluated using MATLAB script, may be different from C/C++ code

•AGR parameters: m=400, s=5

* SVM evaluated using C++ code

5. Information Extraction

5.1 The methods of information extraction

Multiregion image segmentation

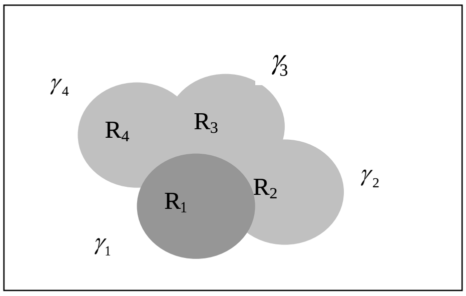
Multiregion Level set method: use multiple level et functions to segment the given image, the region enclosed by each curve is one object feature.

Feature extraction

The ecological feature includes: vegetation, water body, bare soil, and residential areas.

5.2 Multiregion level set method with multiresolution

Representation of a Partition

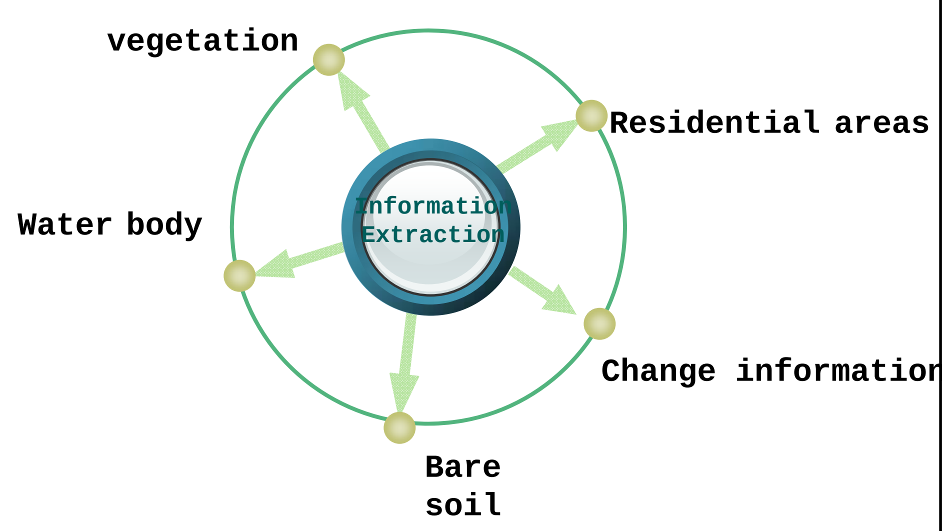


Representation of a partition of the image domain by explicit correspondence between regions of segmentation and regions defined by closed simple planar curves (illustration for five regions).

5.3 Feature extraction

Different operations of Seven-band Landsat TM images can make some characteristics distinct and easily obtain valuable information. Currently, further research put forward a number of prominent favor certain types of index, which is the application of more vegetation index, water index, bare soil index, and residential areas index.

In the ecological feature extraction for the Landsat TM remote sensing data, some features will be strengthened by seven bands of different operations.



Vegetation extraction

- The Normalized Difference Vegetation Index (NDVI) is a important indicator that can be used to analyze remote sensing measurements, typically but not necessarily from a space platform, and assess whether the target being observed contains live green vegetation or not.

$$NDVI = \frac{TM4 - TM3}{TM4 + TM3}$$

Water body extraction

- There are many methods to extract the water body in the multi-band remote sensing image, e.g. Normal Difference Water Index (NDWI), Region Water-contained Index (RWCI), the second-order PCA, humidity feature image in K-T transformation.

The NDWI is defined as:

$$NDWI = \frac{TM2 - TM4}{TM2 + TM4}$$

The RWCI is defined as:

$$RWCI = TM4 + TM5 + TM6 + TM7 - TM1 - TM2$$

Bare soil extraction

- The Normal Different Building Index (NDBI) is used in the bare soil extraction and it is defined as:

$$NDBI = \frac{TM5 - TM4}{TM5 + TM4}$$

The NDBI is used to detect the spatial distribution of urban building area. Combing the NDVI, it can efficiently detect the building and bare soil.

Residential area extraction

- The residential area extraction is realized combining many features. Here we introduce the procedure of its method as follows:
- ***Step 1:*** Use the **NDBI** to obtain the rough location of the residential area

$$\mathrm{Re}\,sident = (NDVI < T1)\,\&\,\&\,NDVI > T2$$

- Where $T1$ is the threshold of the vegetation extraction; $T2$ is the minimal intensity of the residential area.

5.4 Experimental results

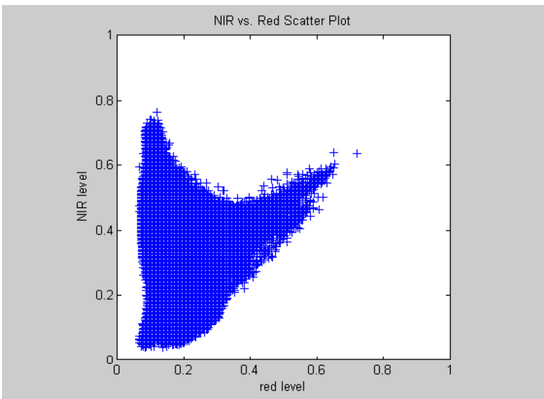
In our experiment, the images are from Wujiaqu City of Changji, Xinjiang in the type of TM.Oribit number is PATH=142, ROW=029. The latitude and longitude are 44.13'N and 87.16'E, respectively. The size is 900*1000. The following two images are taken in 2002 and 2007 by synthesizing band-7, band-4, band-3, respectively.

The PCA analysis

Type	TM1	TM2	TM3	TM4	TM5	TM6	TM7	Eigenvalue
PC1	0.3647	0.1661	-0.46185	-0.15572	0.6881	0.14833	0.32611	74721
PC2	0.2546	0.11284	-0.36723	-0.043655	0.022745	-0.11606	-0.87844	10545
PC3	0.4191	0.19826	-0.41446	-0.037644	-0.66483	-0.23806	0.33636	1497
PC4	0.0176	-0.9107	-0.3415	-0.1380	-0.0912	0.1624	0.01377	350
PC5	0.6012	-0.2968	0.4498	0.18441	0.1996	-0.52758	0.01585	133
PC6	0.4961	0.03291	0.26682	0.20661	-0.1740	0.7759	-0.0808	96
PC7	0.1359	0.04208	0.30736	-0.93633	-0.07444	0.03370	-0.04352	25

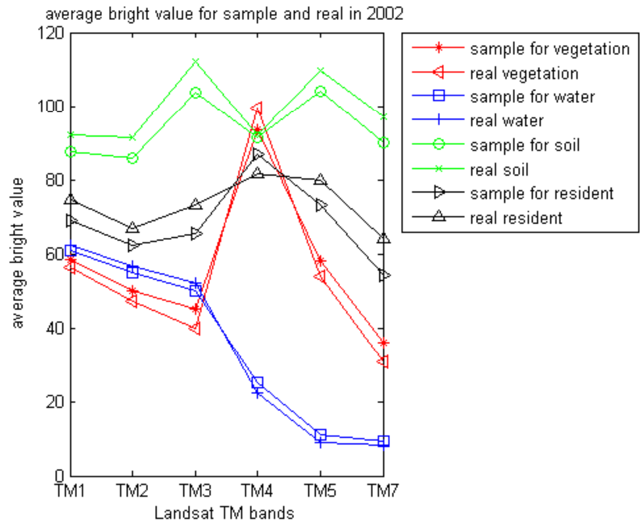
e.g. PCA feature vector matrix for bands in synthetic image 2007

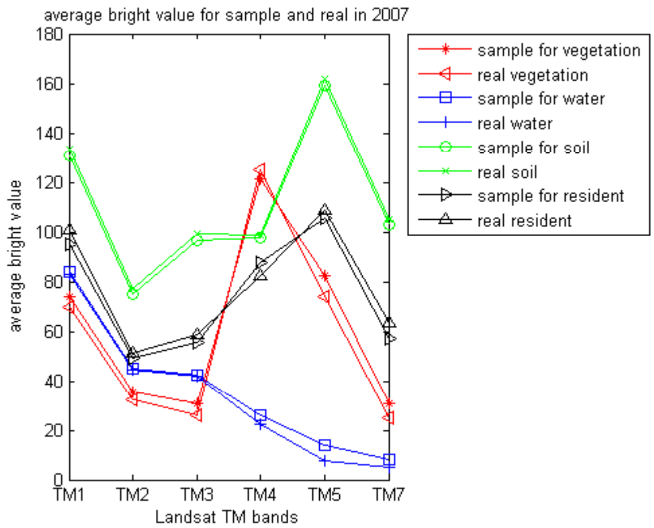
Vegetation extraction



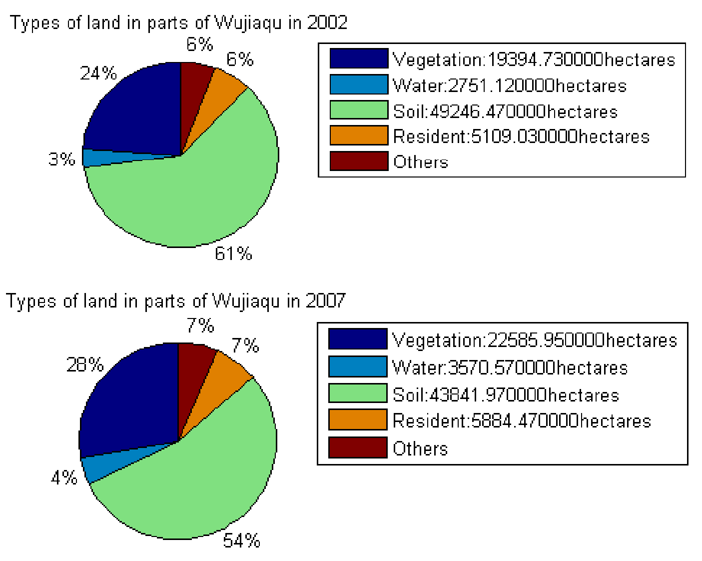
Scatter for near infrared and infrared spectroscopy (x—the value of infrared, y—the value of near infrared)

Change information analysis

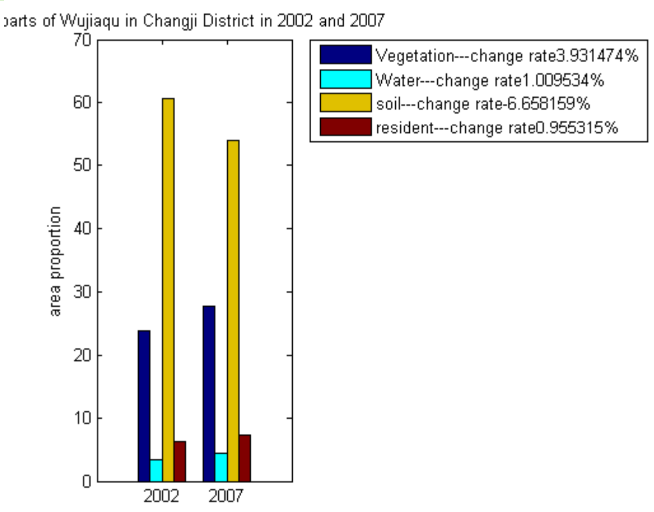




Spectral signatures of four types for study area in 2002 and 2007



Proportion of four types for study area in 2002 and 2007



Change rate of four types for study area between 2002 and 2007

- From the results of change detection, we can conclude that:
- Moisture content and the area of water body are decreasing.
- Especially in the region with more water, the area of residential area and building region are increasing more quickly.
- The area of vegetation is increasing in the north and more bare soil becomes vegetation because many people migrate.

6. Conclusion

Work conclusion

- At the current, three main aspects of our work done are change detection, image classification, and information extraction:
- Change detection: a new method of change detection is proposed. Firstly, the two eigenmatrix are obtained by transforming different synthesis images from different multi-band images using PCA method; Secondly, generalized PCA method is used to transform the changed transformed results convert the image space; Finally, the change object can be detected using the segmentation method based on OTSU.

- Image classification: To analyze the remote sensing images using three segmentation methods, i. e. supervised method, unsupervised method, and semi-Supervised method. At first, we use supervised and unsupervised methods to classify the remote sensing image, such as Maximum Likelihood, SVM, ISODATA algorithms. Then we proposed graph-based semi-Supervised method, which uses a subset of data to design the graph and learn the function, then make prediction on full dataset by weighted combination of anchors.

- **Information extraction:** Firstly, multi-region level set method is proposed to partition the remote sensing image into N homogeneous regions with N-1 level set curves and each region presents one object. Secondly, a number of prominent favor certain types of index, which is the application of more vegetation index, water index, bare soil index, and residential areas index, is used to extract ecological features by different operations of Seven-band Landsat TM images. In the ecological feature extraction, many features will be strengthened by seven bands of different operations.

Work prospect

- We have realized the factions of change detection, image classification, and information extraction. But a lot of work must be done to improve the performance.
- We have made plans to do as the following:
- **Change detection:** improved algorithms e.g.. Level set method, image fusion are employed.
- **Image classification:** improved SVM, Graph-based learning, Mean-Shift.
- **Information extraction:** many features can be extract, e.g. temperature, rainfall. What's more, fusing many features can predict the happenings.

Thank You !

Institute of Image Processing & Pattern Recog

Report on current status of tripartite project in Xinjiang University

Prof. Z. Jia

新疆大学 Xin Jiang University

Report on current status of tripartite project in Xinjiang University

新疆大学 Xin Jiang University

Brief Introduction to

School of Information science and Engineering of Xinjiang University

新疆大学
Xin Jiang University

School of Information Science and Engineering of Xinjiang University was founded in May 2001.

新疆大学
Xin Jiang University

The school now consists of 3 departments:

- Department of Telecommunication Engineering,
- Department of Computer Science and Technology,
- Department of Network and Information Security

新疆大学 Xin Jiang University

Attached research institutes and centers:

- International Institute for scalable software,
- R & D center of multilingual information.

新疆大学 Xin Jiang University

Labs in School

- Multilingual Information Technology Laboratory
- Intelligent Information Technology Laboratory

新疆大学
Xin Jiang University

Major research areas:

- Optical communication and optical sensor theory and technology
- Voice information processing technology
- Digital image information technology
- Modern communications and network technology
- Multi-language information processing

新疆大学
Xin Jiang University

Computer Science and Technology disciplines are capable of granting doctor’s degrees, and with post-doctoral work stations.

新疆大学 Xin Jiang University

Undergraduate specialties

- Computer Science and Technology,
- Network and Information Security,
- Communication Engineering,
- Electronic Information Engineering.

新疆大学 Xin Jiang University

Teaching faculties in School

More than 150 teaching faculties, including:

- 11 professors,
- 44 associate professors,
- 3 senior engineers.

新疆大学
Xin Jiang University

Students in School

The school has more than 1700 students, including:

- 8 PhD students,
- 210 Master students,
- 1500 undergraduate students.

The School enrolls on-job social workers with undergraduate degrees to pursue master's degree in engineering.

新疆大学
Xin Jiang University

The progress

of the 3-party project in Xinjiang University .

新疆大学
Xin Jiang University

The progress of our research

新疆大学
Xin Jiang University

Research Projects

11 postgraduates have participated in related research projects:
——————International Cooperative Research Project of the Ministry of Science and Technology of the P.R.China
(Grant number: 2009DFA12870)

新疆大学
Xin Jiang University

Research Projects

Segmentation and change detection in remote sensing images

———New algorithms

新疆大学
Xin Jiang University

Publication

Journal Articles:

1. Han QingSong, Jia Zhenhong et al., “Remote-sensing image thresholding segmentation based on the modified Otsu algorithm,” Laser Joural (in Chinese), Vol.31, No.6, 33-34, 2010.

2. Lu Binbin, Jia Zhenhong et al., “A new FCM algorithm based on Monkey-king Genetic Algorithm for remote sensing image segmengtation,” Laser Joural (in Chinese), Vol.31, No.6, 15-17,2010.

新疆大学
Xin Jiang University

Publication

3. Huang Ningning, Jia Zhenhong et al., "A new algorithm for remote sensing image segmentation based on the combination of the improved fuzzy entropy and local information," Laser Joural (in Chinese), Vol.31, No.6, 20-22, 2010.

4. Yu Yinfeng, Jia Zhenhong et al., "Change detection in muti-temporal satellite images," Laser Joural (in Chinese), Vol.31, No.6, 25-27, 2010.

5. Liu Wnjing, Jia Zhenhong et al., "A New Noisy Remote Image OTSU Segmentation Algorithm Research," Laser Joural (in Chinese), Vol.31, No.6, 28-30, 2010.

新疆大学
Xin Jiang University

Publication

Journal Articles(In Press):

1. HUANG Ningning, JIA Zhenhong et al., "An Algorithm for Image segmentation Based on the Combination of the Improved FCM and local Information," Computer Applications and Software (in Chinese), 2011.

2. HUANG Ningning, JIA Zhenhong et al., "Segmentation of Textures Based on Gray-Level Co-occurrence Matrix Features and Fast EM Algorithm," Communication Technology (in Chinese), 2011.

3. HUANG Ningning, JIA Zhenhong et al., "An Algorithm for Remote Sensing Image Segmentation Based on the Combination of the Improved FKCN and local Information," Computer Engineering and Application (in Chinese), 2011.

新疆大學 Xin Jiang University

Publication

Journal Articles(In Press):

4. YU Yin-feng, JIA Zhen-hong et al., "An Unsupervised Change Detection Algorithm in Satellite Images Based On NonSubsampled Contourlet Transform and Pulse Coupled Neural Network," Computer Engineering and Application (in Chinese), 2011.

5. WANG Yalan, JIA Zhenhong et al., "Remote sensing image denoising based on NSCT and PCA," Computer Engineering and Application (in Chinese), 2011.

6. Liaoyan Jia Zhenghong, " Based on micro canonical annealing and improved FCM algorithm for remote sensing image segmentation," Laser Joural (in Chinese) ,2011.

新疆大學 Xin Jiang University

Publication

Journal Articles(In Press):

7. YU Yin-feng, JIA Zhen-hong et al., "An Unsupervised Change Detection Algorithm in Satellite Images Based On Principal Component Analysis and Up-Down-Set Fuzzy Kohonen Clustering Network," Journal of Optoelectronics.Laser (in Chinese), 2011.

8. Han QingSong, Jia Zhenhong et al., "Remote sensing image thresholding segmentation based on the local spatial information," Journal of Computer Applications(in Chinese), 2011.

9. ZHAO Lei, Zhenhong Jia "Optimization of RBF Network based on k-means and Quantum Genetic Algorithm," Communication Technology (in Chinese), 2011.

新疆大学
Xin Jiang University

Publication

Conference Proceedings Articles:

1. Yingjie Gu, ZhenhongJia et al., “A modified fuzzy C-means with particle swarm optimization adaptive image segmentation algorithm,” 2010 International Conference on Display and Photonics (ICDP2010).

2. Han QingSong, Jia Zhenhong et al., “Handwritten Digits Recognition Using HMM and PSO based on storks,” 2010 International Conference on Display and Photonics (ICDP2010).

新疆大学
Xin Jiang University

Patent Submission

Liu Wnjing, Jia Zhenhong et al., “An Improved 2D Otsu method for Noisy Remote Image Segmentation,” China Patent, No.201010243654.9

Xin Jiang University

Examples of research results in project

Example 1

A modified Otsu algorithm for image Segmentation

Xin Jiang University

Examples of research results in project

Results Comparison :

A. Traditional Otsu algorithm

B. Other modified Otsu algorithm

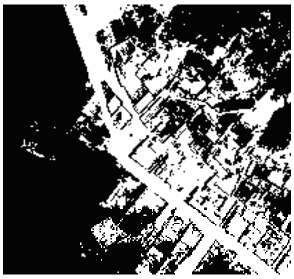
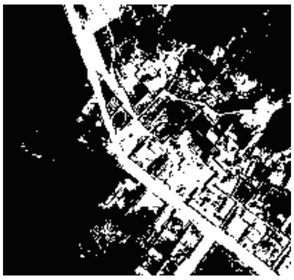
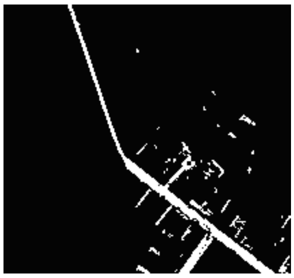
——— D.Y.Huang, et al., Optimal multi-level thresholding using a two-stage Otsu optimization approach, Pattern Recognition Letters, 30 : 275－284, 2009.

C. Our modified Otsu algorithm

Xin Jiang University

Examples of research results in project

Results Comparison :

A

C

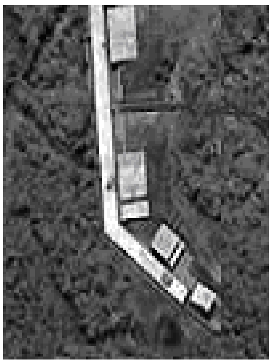
0.029s 1.311s

Xin Jiang University

Examples of research results in project

Example 2

Image Segmentation Algorithm Based on the Otsu’ s Rule and K-means Clustering




Examples of research results in project

Results Comparison :

A. Traditional Otsu Algorithm

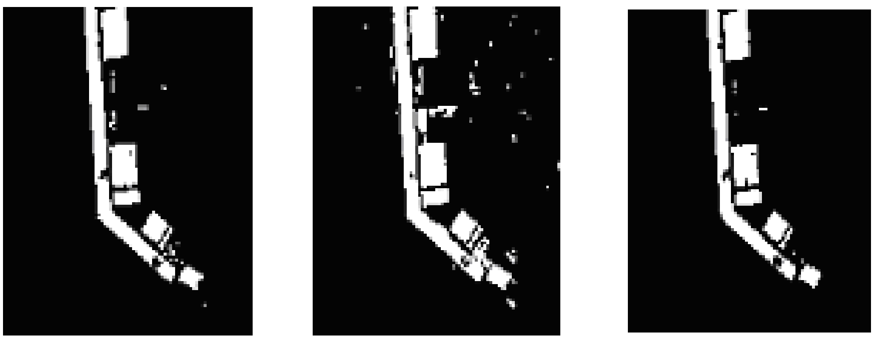
B. An Improved K-means Algorithm
——— MIN Li et al., Improved K-means Algorithm For Medical MR Images Segmentation[J]. IEEE Computer Science, 2009, 118(2):285-288.

C. Our Algorithm

新疆大学
Xin Jiang University

Examples of research results in project

Results Comparison :



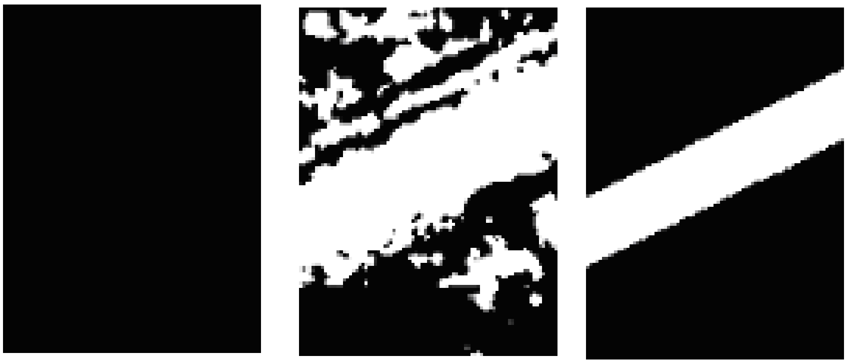
A B
C

13.720s 0.159s

Xin Jiang University

Examples of research results in project

Results Comparison :



A B

C

14.622s 0.171s

0.068s

Xin Jiang University

Academic exchange and cooperation

新疆大学 Xin Jiang University

"Introducing talents cultivation project"
"千人计划引智培育项目"

——— a cultivation program for

"Thousand Talents Plan "

"千人计划"

新疆大学 Xin Jiang University

"Introducing talents cultivation project"
"千人计划引智培育项目"

- Support overseas outstanding youth to short-term work in China,

- Foster them to be "Thousand Talents Plan" reserve candidates.

新疆大学
Xin Jiang University

Preparations for the establishment of joint research centre in Urumqi

(ICIIPEP)

新疆大学
Xin Jiang University

Preparations for ICIIPEP

We are ready for

- 100 square meters of space,
- 10 computers,
- 2 servers,
- Some research software.

新疆大学
Xin Jiang University

Preparations for ICIIPEP

We are seeking other funding from Xinjiang University.

新疆大学
Xin Jiang University

Application for other projects

新疆大学
Xin Jiang University

To promote scientific cooperation with American and Oceanian region and high-level personnel training programs

"促进与美大地区科研合作与高层次人才培养项目 "

——— Supported by Department of International

Cooperation and Exchange,

the Ministry of Education of the People's Republic of

China

新疆大学
Xin Jiang University

Intelligent remote sensing digital image processing technology and its application

智能化遥感数字图像处理技术及应用研究

———Project is approved,

Thank you！

Mining High Speed Data Streams

Dr. Russel Pears

Tripartite Workshop February 2011

Mining High Speed Data Streams

Russel Pears

Overview

- Short Biography
- My Research interests in Data Mining
- Challenges in Mining Concept Drifting Data Streams
- A concept-based approach: CBDT
    - Algorithm
    - Experiments
    - Results & Analysis

Short Bio

- Currently Senior Lecturer in Computer Science and Programme Leader for the PhD Programme in the school of Computing
- Main research interests are in the areas of Data Mining and Machine Learning.
- Currently involved in projects on Mining Data Streams and Automatic Rule Deduction from large datasets

Major Publications in Recent Past

1. Dynamic Interaction Networks Versus Local Trend Models For Multiple Time Series Prediction, Journal of Cybernetics and Systems, to appear in February 2011.
2. Automatic Item Weight Generation for Pattern Mining and its Application, accepted for publication in the International Journal for Data Warehousing and Data Mining, to appear in March 2011.
3. Multiple Time-series Prediction Through Multiple Time-series Relationships Profiling and Clustered Recurring Trends, accepted for publication at the 15th Pacific-Asia Conference in Knowledge Discovery and Data Mining to be held in May 2011
4. CBDT: A Concept Based Data Stream Miner in Proceedings of the 13th Pacific-Asia Conference Knowledge Discovery and Data Mining, 2009

The problem of concept drift

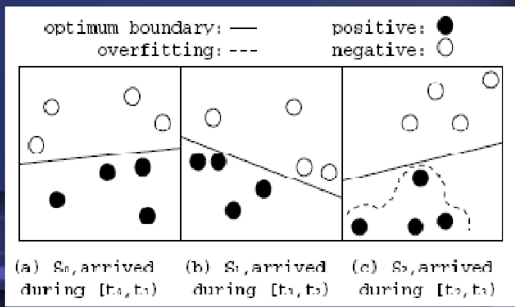
- Every machine learning method looks to create a concept description of the underlying data set
- Classical data mining uses a static (historical) data set to induce a model which is used for future prediction
- Underlying assumption here is that data follows a stationary stochastic distribution
- However, most real-world data sets have a changing (=drifting) data generation process (seasons, etc.)
- Thus concepts (patterns) learnt from period p_1 does not hold in period p_2
- This requires an incremental approach to learning that will continuously refine and update concepts learnt in the past

High-Speed Data Stream Mining

- Increasingly, massive amounts of data are being generated by applications such as : e-commerce, telecommunications, scientific projects like the SKA telescope
- This data might contain valuable knowledge, but its sheer rate of generation outpaces the speed at which the data can be mined
- Sometimes, volume and data creation rates present a challenge for possible storage solutions
- Such volume-intensive data streams cannot be mined by traditional algorithms

Problems with using Sliding Time Windows to build Models

- The standard method of controlling memory requirements is a time sliding window
- Possible loss of valuable information as data slides out of the window:



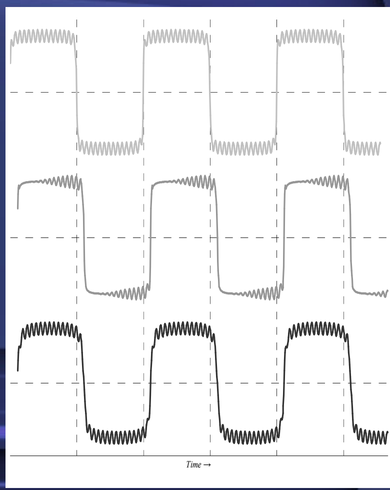
- S_0 is similar to S_2
- S_1 removes S_0 as data slides out of the window
- S_2 needs to re-learn

A concept-based approach

- Main objective: Remove the arbitrarily sized time window to retain data about the data stream's history
- Time-window based algorithms remove information determined by $f(time)$
- The concept-based approach removes information by $f_{min}(information\text{-}loss)$ – the *content* of the data stream decides which information is out-dated

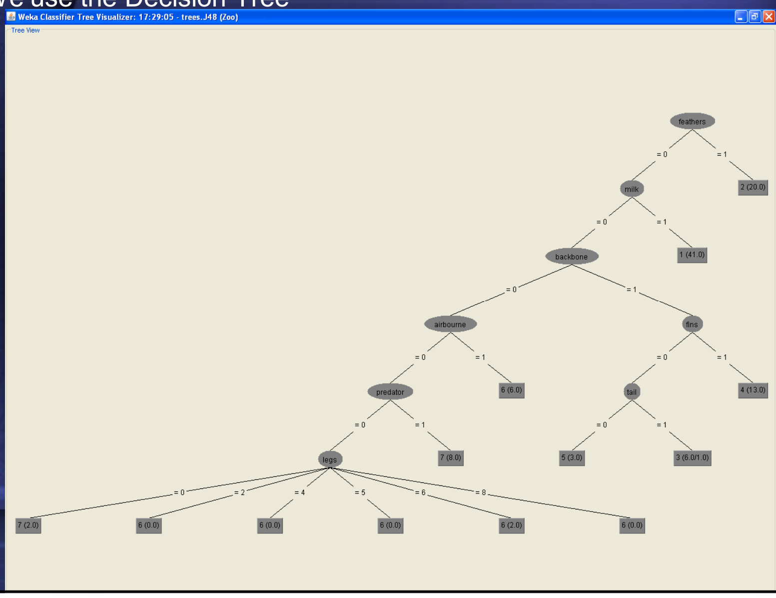
A concept-based approach – cont

- Top: super-imposed concept drift in a data stream
- Middle: time-window based approach needs to re-learn super-imposed concept every time the main concept changes
- Bottom: concept-based approach still remembers super-imposed drift



Basic Mining Approach

- We use the Decision Tree

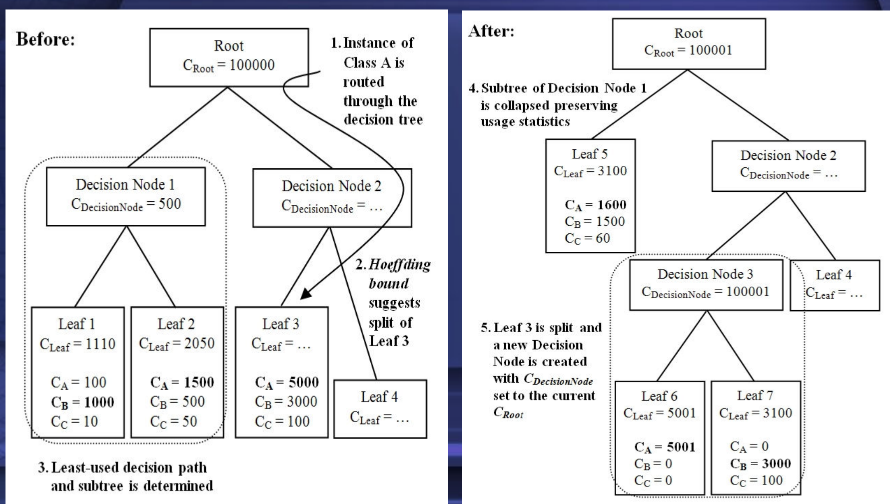


CBDT Learner

- The learner maintains a forest of trees with each tree rooted on a different attribute - this allows to be more responsive to abrupt changes in the concept
- A purge control mechanism decides which information is out-dated
- A memory allocator forces the individual trees to fight for resources, as it allocates more memory to highly accurate trees

CBDT – A concept-based approach to High-speed Data Stream Mining

The basic algorithm



The basic algorithm – memory

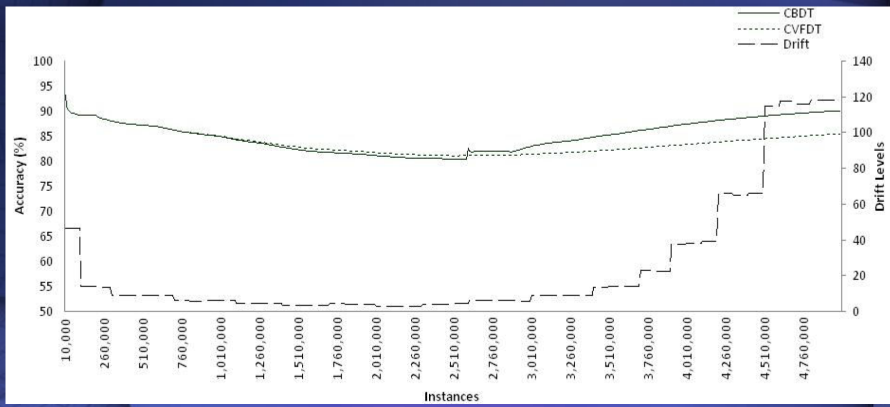
- Each tree in the forest grows independently and competes with other trees for memory of the total available pool
- Each time the memory allocator is invoked, it allocates the available memory proportional to each tree's current classification accuracy
- By using a logarithmic function, we bias memory allocation to high performance trees, thus limiting sensitivity to highly erratic concept drifts

Experiments

- A synthetic data generator was used to simulate hyperplanes with concept drift in a controlled manner
- 2 classes (A and B)
- 25 dimensions each taking one of 5 discrete values
- 5,000,000 instances
- Noise was added by randomly switching the class with 5% of the generated instances
- Compared classification accuracy of CVFDT and CBDT for various drift types

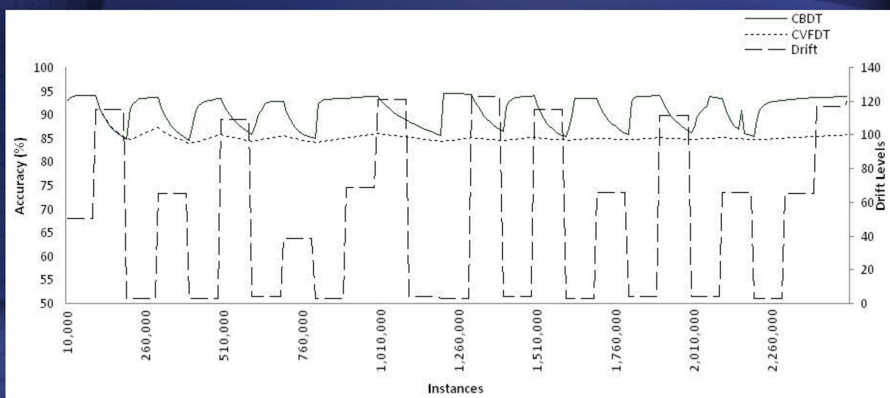
CBDT – A concept-based approach to High-speed Data Stream Mining

Results – Smooth Drift



CBDT – A concept-based approach to High-speed Data Stream Mining

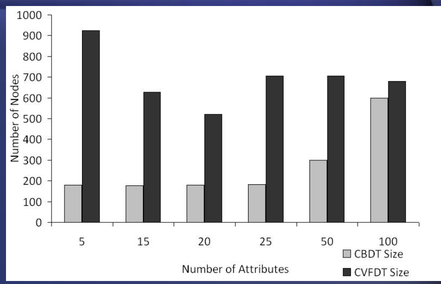
Results – Dramatic Drift

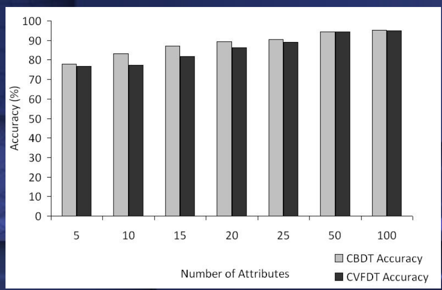


CBDT – A concept-based approach to High-speed Data Stream Mining

Results – Scalability

Memory consumption by data dimensionality

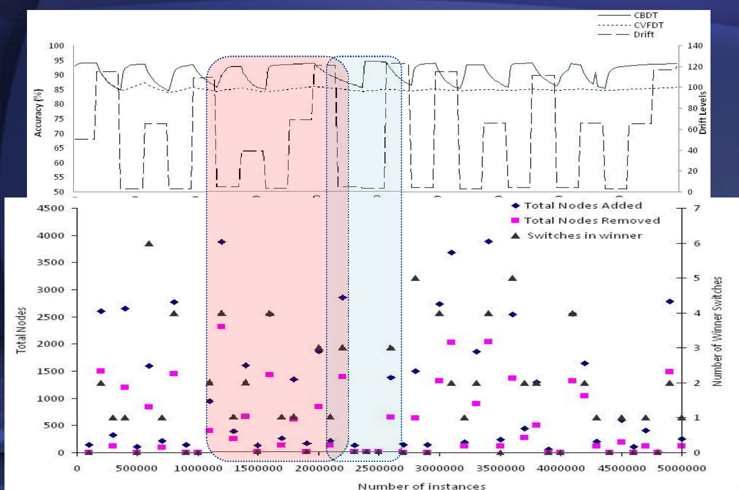




Classification accuracy by data dimensionality

CBDT – A concept-based approach to High-speed Data Stream Mining

Results – Internal dynamics



Analysis

- Slow drift: Both algorithms exhibit similar behaviour at low drift levels, but with increasing rate of concept drift CBDT gets more accurate
- Dramatic drift: While CVFDT is unable to react to changes, CBDT adjusts very well, its accuracy being well synchronised with the drift
- Accuracy vs Dimensionality: Both algorithms scale similarly, CBDT slightly better, but less with increasing dimensionality
- Memory vs Dimensionality: CBDT performs much better throughout due to the flexibility of the forest

Conclusion

- Concept-based approaches like CBDT can out-perform time-window based approaches such as CFVDT on non-specialised data sets
- CBDT maintains even a higher level of accuracy than CVDFT while using lesser memory
- CBDT can retain a high level of accuracy even in the face of massive changes in concept drift

CBDT – A concept-based approach to High-speed Data Stream Mining

Future work

- Since each tree in CBDT's forest is grown independently, it lends itself to a high degree of parallelisation – which is worth being investigated, especially in respect to issues with the centralised memory allocation routine
- Another interesting topic would be to devise a metric for quantifying the importance of recurring pattern – this would improve long term memory and allow the algorithm to proactively forget patterns with low significance

CBDT – A concept-based approach to High-speed Data Stream Mining

References

Hulten, G., Spencer, L. & Domingos, P. (2001). Mining time-changing data streams. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 97-106

Domingos, P. & Hulten, G. (2000). Mining high-speed data streams. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, 71-80.

Guah, S., Meyerson, A., Mishra, N., Motwani, R. & O'Callaghan, L. (2003). Clustering Data Streams: Theory and Practice. TKDE Special Issue on Clustering, 15.

Hulten, G., Spencer, L. & Domingos, P. (2001). Mining time-changing data streams. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 97-106.

Papadimitriou, S., Brockwell, A. & Faloutsos, C. (2004). Adaptive, unsupervised stream mining. The VLDB Journal, 13(3), 222-239.

Aggarwal, C. C., Han, J., Wang, J. & Yu, P. S. (2004). On demand classification of data streams. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 503-508.

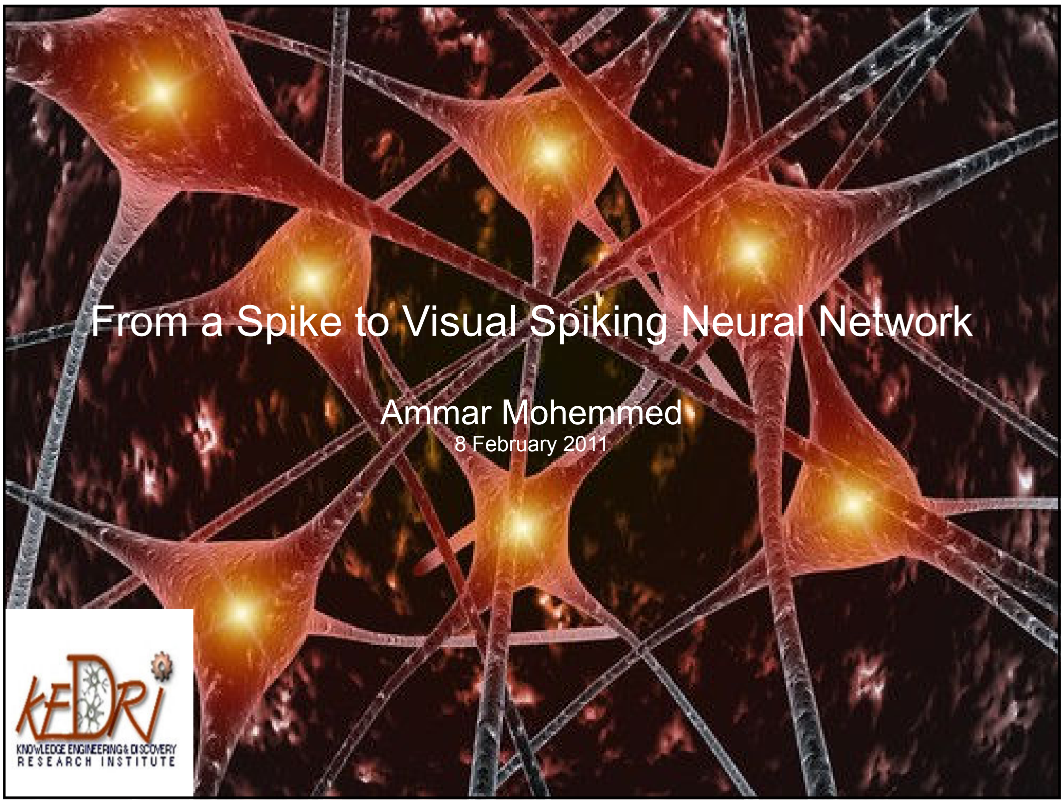
Wang, H., Fan, W., Yu, P. S. & Han, J. (2003). Mining concept-drifting data streams using ensemble classifiers. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 226-235

Questions

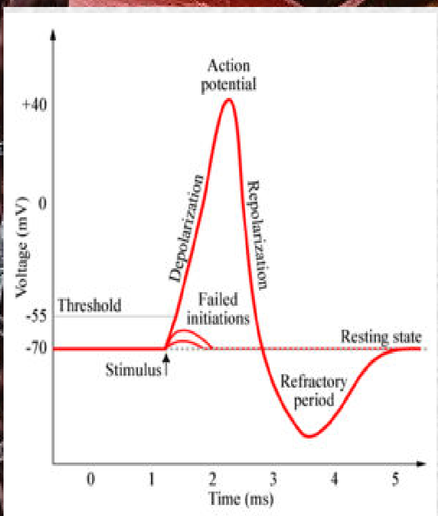
Thank you.

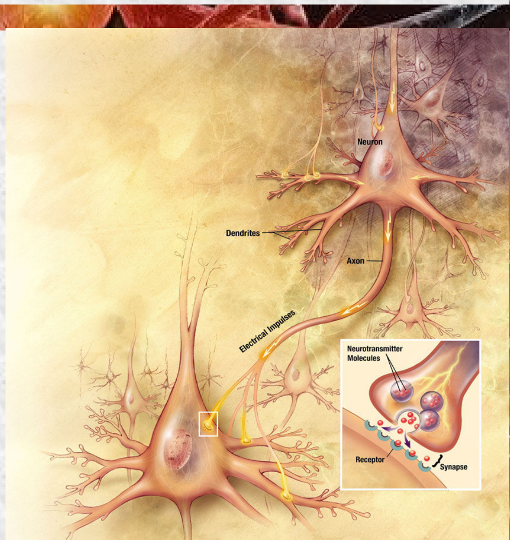
From a Spike to Visual Spiking Neural Network

Dr. Ammar Mohemmed



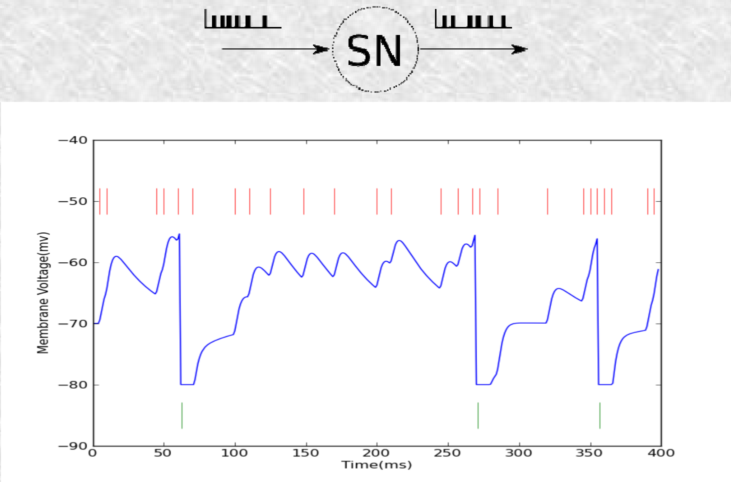
What is a spike ?





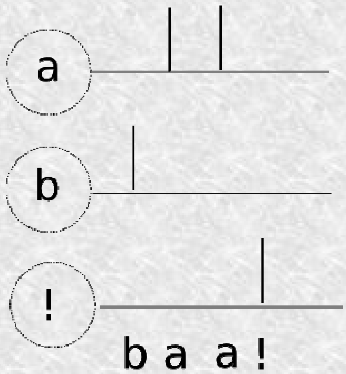
source: Wikipedia

Spiking Neuron

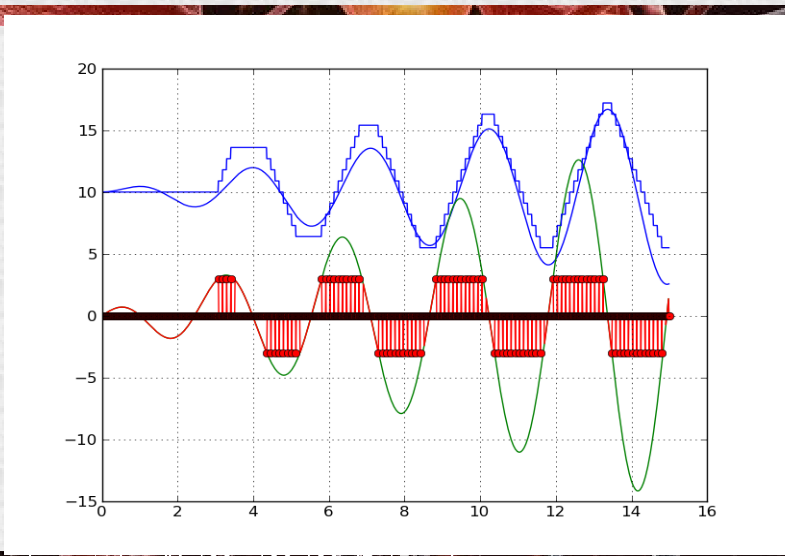


How information is encoded in SNN ?

- Rate encoding, Rank encoding, Temporal encoding

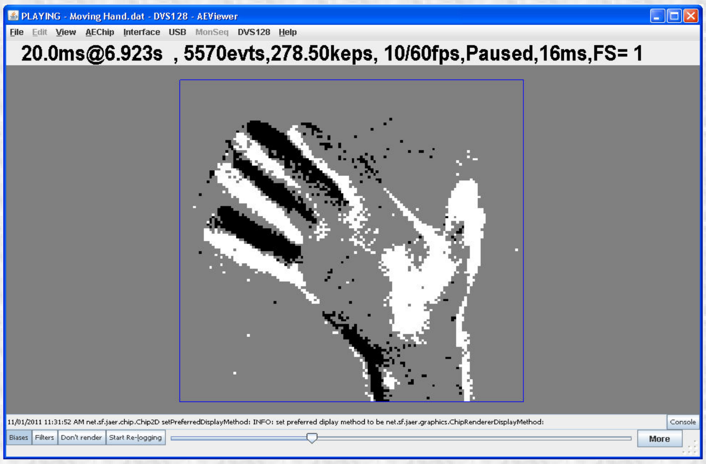


How information is encoded in SNN ?



Lichtsteiner, et al .; , "A 128× 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," Solid-State Circuits, IEEE Journal of , Feb. 2008

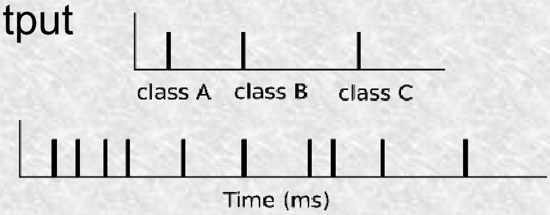
How information is encoded in SNN ?

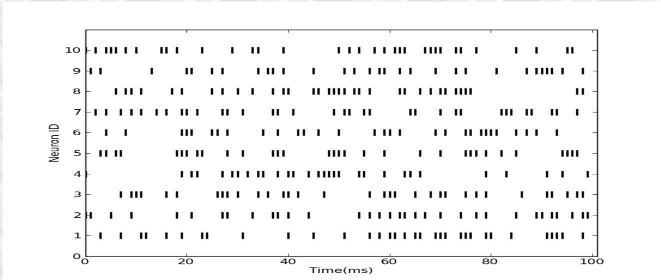


Institute of Neuroinformatics, ETH/UNI Zurich

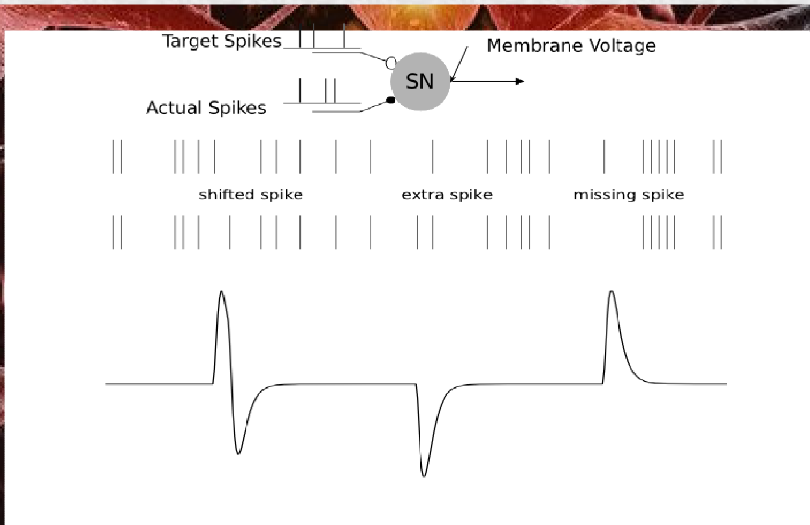
How information is decoded in SNN ?

- SNN output



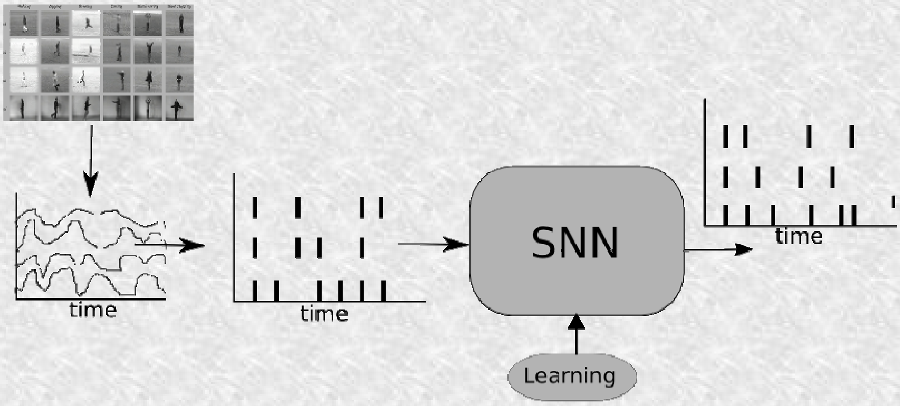


A Spiking Neuron as a Similarity Measure



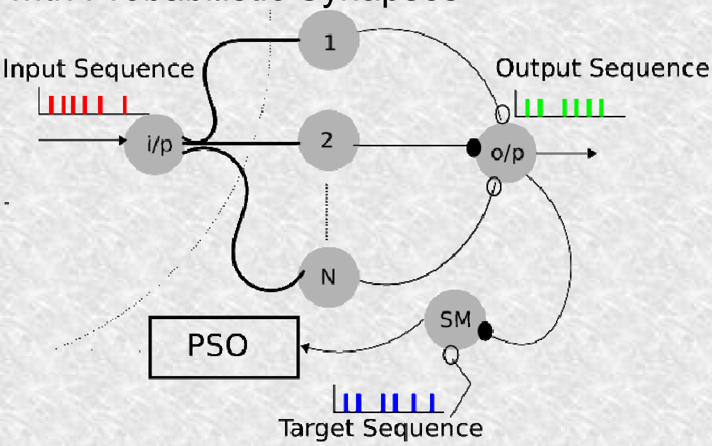
A Spiking Video System

- Human Activity Detection

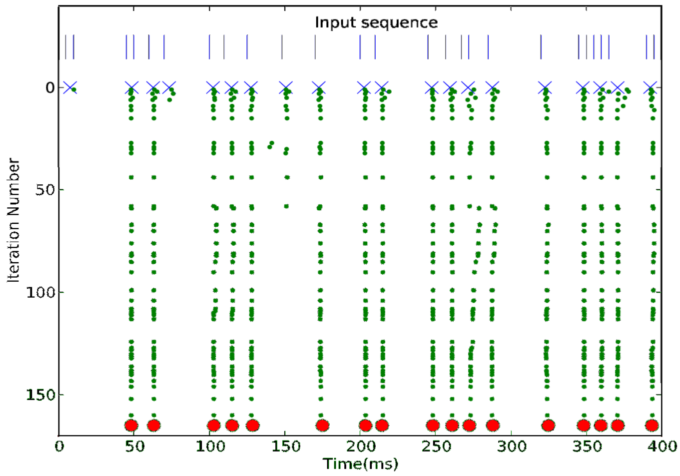


Sequence Learning

- SNN with Probabilistic Synapses



Example Output



Conclusion

- SNN for large number of classes
- Time is well represented in SNN
- Spiking Neural Network is naturally more adequate for spatio-temporal data
- SNN might simplify computation
- SNN for Remote Sensing ?!

Multiple Time-Series Prediction for Environmental Decision Support

Harya Widiputra

MULTIPLE TIME-SERIES PREDICTION FOR ENVIRONMENTAL DECISION SUPPORT

Harya Widiputra
The Knowledge Engineering and Discovery Research Institute
Auckland University of Technology

KNOWLEDGE ENGINEERING & DISCOVERY RESEARCH INSTITUTE

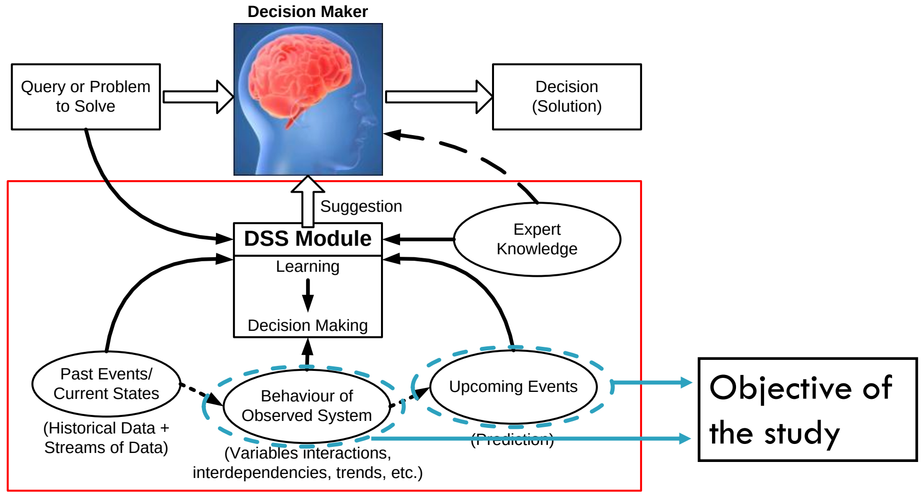
Auckland, 8 February 2011

AUT UNIVERSITY

Presentation Outline

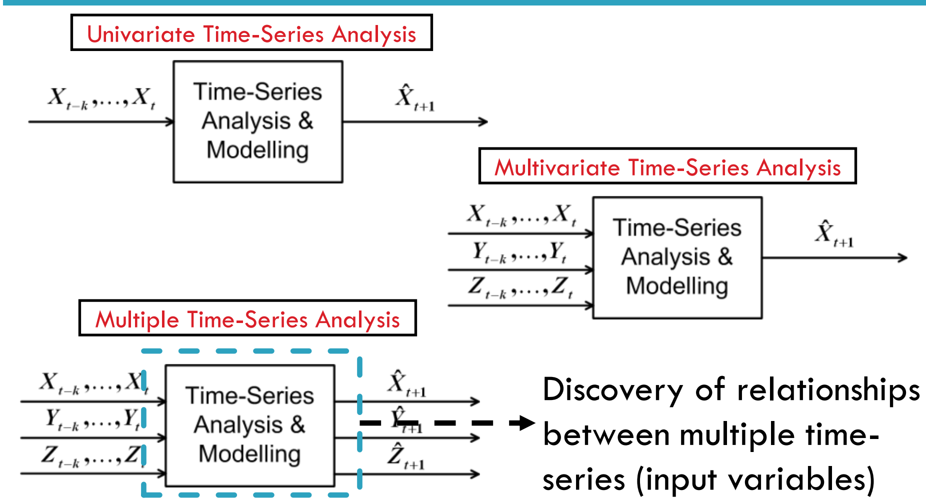
1. Background
2. Method of Time-Series Analysis & Modelling
   a. Methods of Time-Series Analysis
   b. Local Model
3. Multiple Time-Series Analysis & Modelling
   a. Localized Trends Model
   b. Predicting Air Pressure Level in New Zealand
   c. Multiple Time-Series Analysis for Decision Support System
4. Summary & Future Direction

1. Background



http://www.kedri.info harya.widiputra@aut.ac.nz AUT 2011

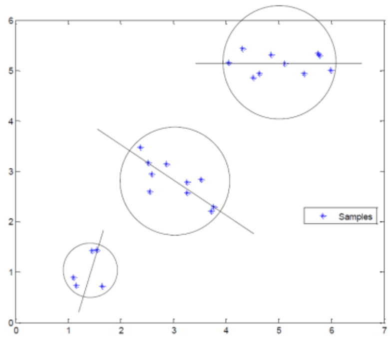
2.a. Methods of Time-Series Analysis



http://www.kedri.info harya.widiputra@aut.ac.nz AUT 2011

2.b. Local Model

- Local models is constructed by grouping together data that has similar behaviour.
- Often based on clustering techniques.
- As different types of phenomena will define their own clusters, specific models are then developed for each cluster.



- Having a set of local models also offers greater flexibility as predictions can be made either on the basis of a single model, a number of selected models or on a global level if needed.

3.a. Localized Trends Model

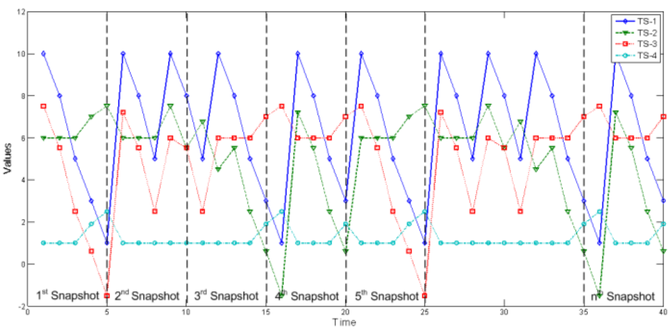
Provides localized profiles of relationships between multiple time-series.

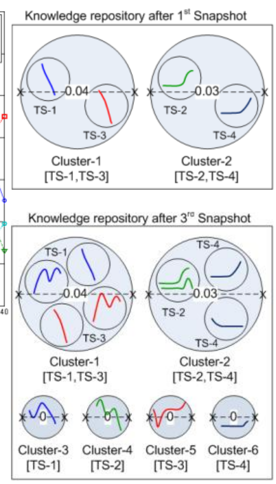
Utilizes a 2-level local modelling process;

1) Relationship profiling between series in a sub-space of the given multiple time-series data.
   - Correlation coefficients with Pearson's correlation
   - Dissimilarity measurement by RNOMC matrix [1]
2) Recurring trends clustering across pairs of time-series in each profile.
   - Trends extraction with non-parametric regression
   - Trends clustering with modified ECM [2]

http://www.kedri.info harya.widiputra@aut.ac.nz AUT 2011

3.a. Localized Trends Model





1st Snapshot

Pearson Correlation

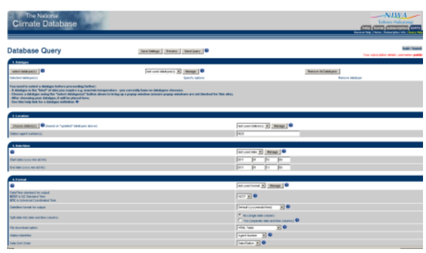
	TS-1	TS-2	TS-3	TS-4
TS-1	1	-0.85	0.96*	-0.85
TS-2	-0.87	1	-0.86	0.97*
TS-3	0.96*	-0.86	1	-0.87
TS-4	-0.85	0.97*	-0.87	1

Normalized Correlation

	TS-1	TS-2	TS-3	TS-4
TS-1	-	-	0.04	-
TS-2	-	-	-	0.03
TS-3	0.04	-	-	-
TS-4	-	0.03	-	-

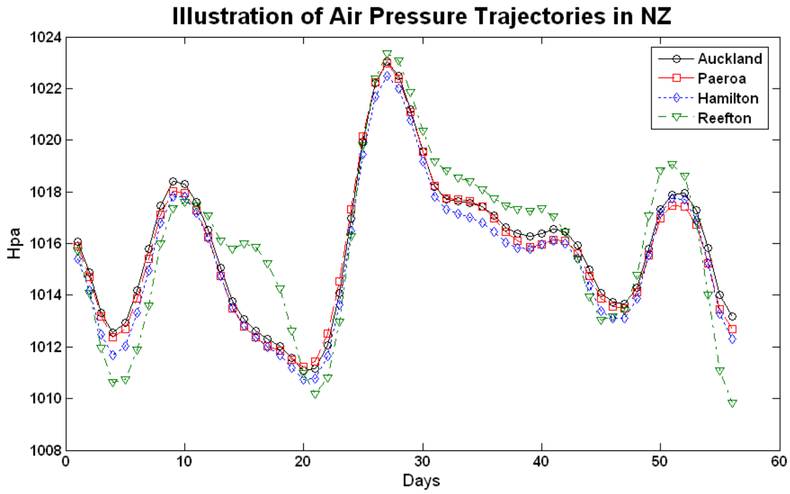
3.b. Prediction of NZ Air Pressure

- Data is acquired from NIWA, the National Institute for Weather and Atmosphere (www.niwa.co.nz)
- Provides historical and current observation of climate and weather conditions in New Zealand.

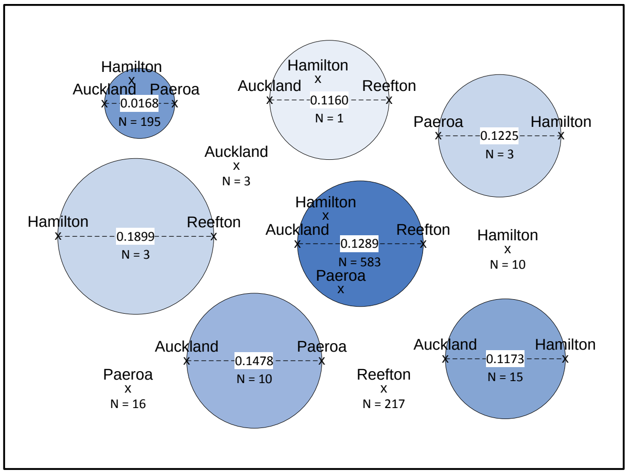


- Covers: temperature, precipitation, wind speed & direction, air pressure level, etc. including pollution level in some area.
- Air pressure levels collected daily in four observation stations: Auckland, Paeroa, Hamilton (North Island, NZ) and Reefton (South Island, NZ) spanning from 1st Jan 2007 to 31st Dec 2009.

3.b. Prediction of NZ Air Pressure

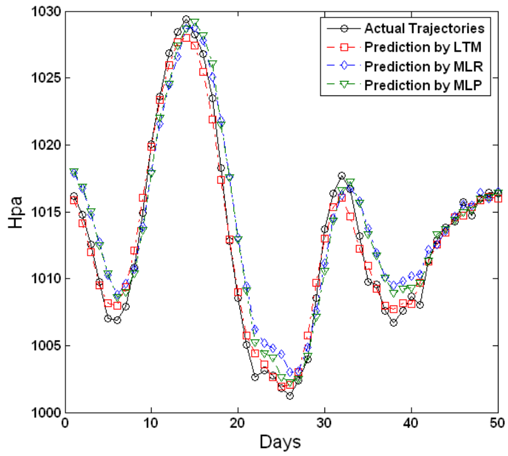


3.b. Prediction of NZ Air Pressure



- Cluster radius represents the farthest distance (*normalized correlation*) between time-series in the same cluster.
- Relative positioning of the labels indicates the degree of similarity in behaviour.

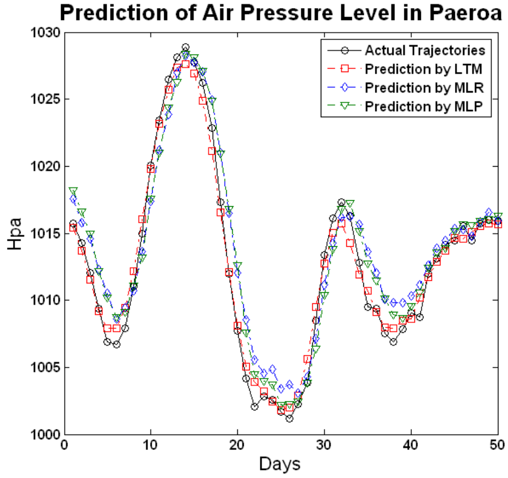
3.b. Prediction of NZ Air Pressure



Prediction Error Rates in RMSE

Location	LTM	MLR	MLP
Auckland	**0.948**	2.194	2.108
Paeroa	0.930	2.173	2.041
Hamilton	0.946	2.275	2.198
Reefton	1.063	2.645	2.554

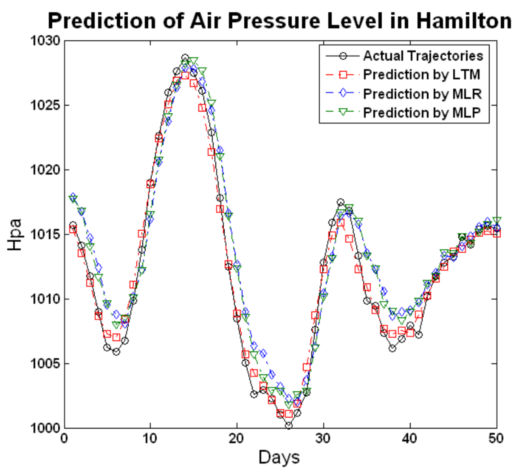
3.b. Prediction of NZ Air Pressure



Prediction Error Rates in RMSE

Location	LTM	MLR	MLP
Auckland	0.948	2.194	2.108
Paeroa	**0.930**	2.173	2.041
Hamilton	0.946	2.275	2.198
Reefton	1.063	2.645	2.554

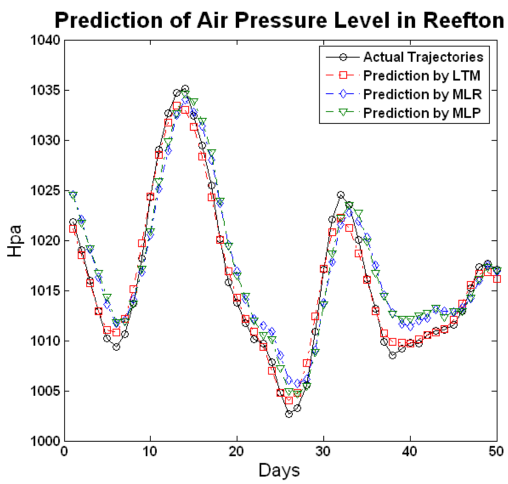
3.b. Prediction of NZ Air Pressure



Prediction Error Rates in RMSE

Location	LTM	MLR	MLP
Auckland	0.948	2.194	2.108
Paeroa	0.930	2.173	2.041
Hamilton	**0.946**	2.275	2.198
Reefton	1.063	2.645	2.554

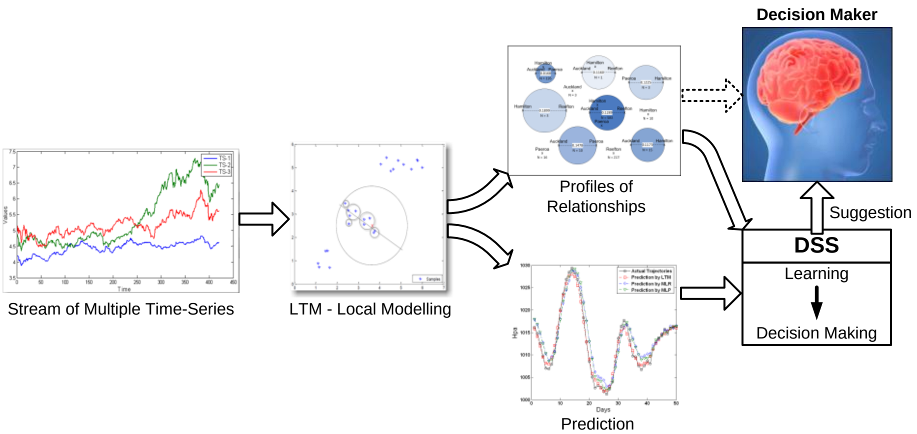
3.b. Prediction of NZ Air Pressure



Prediction Error Rates in RMSE

Location	LTM	MLR	MLP
Auckland	0.948	2.194	2.108
Paeroa	0.930	2.173	2.041
Hamilton	0.946	2.275	2.198
Reefton	**1.063**	2.645	2.554

3.c. Multiple Time-Series Analysis for DSS



4.a. Summary

- LTM demonstrates the ability to extract profiles of relationships and recurring trends from a multiple time-series data.
- Additionally, outcomes of conducted experiments show that predicting movement of multiple time-series using profiles of relationships improves the prediction accuracy.
- This finding leads to the possibility of integrating the LTM with a DSS module.

4.b. Future Direction

- As future work, we plan to explore the use of correlation analysis methods which are capable of detecting non-linear correlations between observed variables (i.e. correlation ratio, Copula, etc.)

Thank You

KNOWLEDGE ENGINEERING & DISCOVERY RESEARCH INSTITUTE

AUT UNIVERSITY

http://www.kedri.info harya.widiputra@aut.ac.nz AUT 2011

References

1. Rodrigues P., J. Gama, and P. Pedroso. Hierarchical Clustering of Time-series Data Streams. IEEE Transaction. Transaction on Knowledge and Data Engineering (2008), Vol. 20, 5:615-627.
2. Song, Q., and Kasabov, N. ECM - A Novel On-line Evolving Clustering Method and Its Applications. In: Posner, M.I. (Ed.), Foundations of Cognitive Science (2001), 631-682.
3. Vitousek, P. M. Beyond Global Warming: Ecology and Global Change. Ecological Society of America. Ecology, Vol. 75 (1994), 7:1861-1876.

Novel Models for Wind Forecasting in the Context of Integrative Decision Support System Framework

Marin Karaivanov

Novel models for wind forecasting in the context of integrative decision support system framework

Tripartite Workshop, 8 Feb 2011, Auckland, New Zealand

Marin Karaivanov, PhD student

Knowledge Engineering and Discovery Research Institute (KEDRI)
(http://www.kedri.info)

Motivation and Driving Factors

- Kyoto protocol for reducing emission of greenhouse gases
- New Zealand commitment to alternative energy sources (90% renewable electricity by 2025) and wind wnergy expansion worldwide
- The need for innovative and accurate methods of wind energy prediction and resource optimization

Wind Energy Worldwide- 2010

Wind Power Worldwide June 2010

Position	Country	Total capacity June 2010 [MW]	Added capacity June 2010 [MW]	Total capacity end 2009 [MW]
1	USA	**36.300**	1.200	35.159
2	China	**33.800**	7.800	26.010
3	Germany	**26.400**	660	25.777
4	Spain	**19.500**	400	19.149
5	India	**12.100**	1.200	10.925
6	Italy	**5.300**	450	4.850
7	France	**5.000**	500	4.521
8	United Kingdom	**4.600**	500	4.092
9	Portugal	**3.800**	230	3.535
10	Denmark	**3.700**	190	3.497
	Rest of the World	**24.500**	2.870	21.698
Total		**175.000**	**16.000**	**159.213**

© WWEA 2010

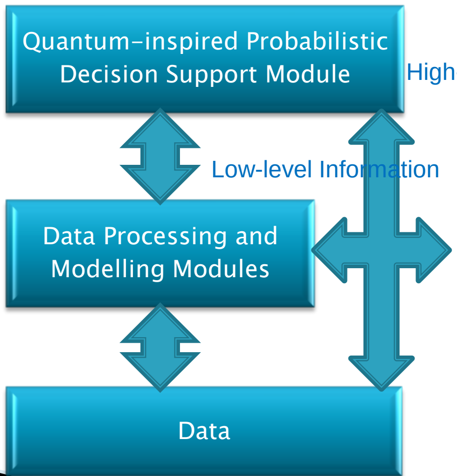
Goals and objectives

- Integrative Decision Support System for Wind Energy Generation that can be successfully deployed in business practice.
- Novel approach for Decision Making, that overcomes the restrictions set by the conventional methods (Quantum- inspired probabilistic Decision Support System, incorporating the principle of superposition)
- Forecasting models which are:

  Robust, producing accurate forecasts
  Adaptive to new data
  Process information in an online mode
  Evolving structure
  Relatively easy to implement

- Apply the proposed framework to Environmental Management, Ecological prognosis and related areas

Integrative Decision Support System Framework



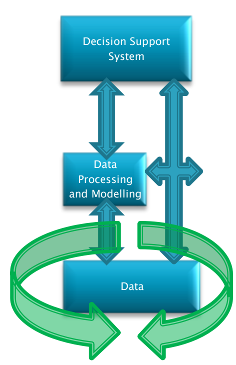
Data structure

- *Time series* – sequence of data points, measured at successive times, spaced at uniform time intervals
- *Time series forecasting* is the use of model to forecast future events based on known past events (e.g. to forecast future data points before they are measured)
- Input data organization

  The sequence of data measurements are organized in vector form

  $X = [\, x(t-4s)\ x(t-3s)\ \ x(t-2s)\ \ x(t-1s)\ \ x(t)\ \ x(t+1s)]\ \ (1)$

  X is the input pattern
  $x(t)$ is the value of time series at the present time t
  $x(t+1s)$ is the point we want to predict
  s is the chosen step

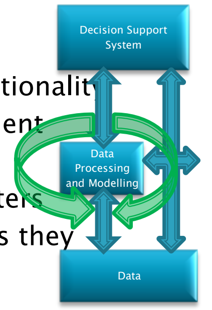


ECOS Framework

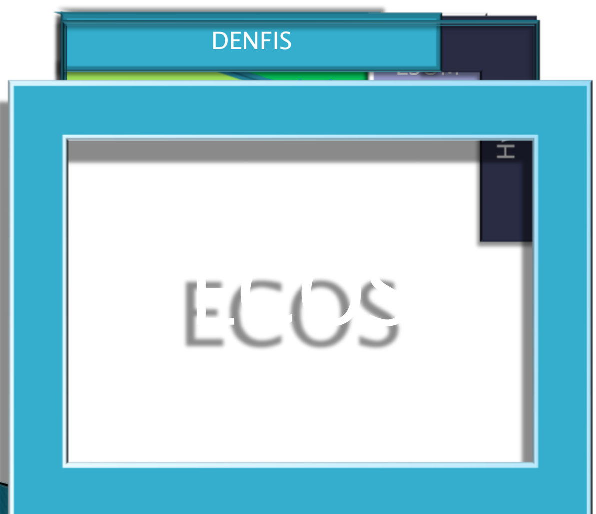
ECOS = Evolving Connectionist Systems

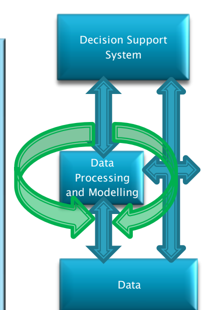
Evolving systems [1] are:

- Systems that evolve their structure and functionality over time through interaction with environment
- Systems that have some predefined parameters (knowledge) but they also learn and adapt as they operate
- Systems that emerge, evolve, develop, unfold through learning and through changing their structure



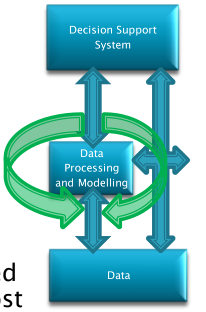
ECOS Framework (cont.)





Dynamic Evolving Neural-Fuzzy Inference System (DENFIS)

- DENFIS [2] incorporates adaptive online and offline learning, oriented for dynamic time series prediction
- DENFIS evolves through incremental, hybrid learning (supervised/ unsupervised)
- DENFIS accommodates new input data, features, classes through local element tuning
- At each moment, the output of DENFIS is calculated through a fuzzy inference system based on *m*-most activated fuzzy rules, dynamically chosen from a fuzzy rule set
- DENFIS employs Takagi-Sugeno-type fuzzy rules



DENFIS applications for energy related prognostic system

- Electricity Grid Optimization

Wind power online forecasting systems

Wind power penetration facilitation and optimization

Electricity demand/ consumption prediction

Environmental and ecological risk prognosis

Wind Forecasting Horizons

- **Dual need for wind forecasts:**
- For integration and <u>trading</u> of power generated by wind sites: 5-12 hours horizon
- For <u>operation</u> and maintenance of wind sites: 0-2 hours horizon

- **To satisfy <u>both</u> of these conditions:**
- 10 min ahead
- Next hour average
- Next 2 hours average
- Next 5 hours average
- Next 12 hours average

Models evaluation

Models evaluation

For evaluation purposes several methods are considered,representing techniques with different learning algorithms and approaches, namely:

- DENFIS
- TWNFI
- KNN-5
- MLP
- MLR

The evaluation is performed on the basis of:

- RMSE and NDEI
- Internal structure and behavior of the model

Experimental Results

As a result from the simulation, 4 models for wind speed prediction were constructed, 1 model for Wind Direction and 1 model for Wind Gust forecasting

Wind Speed prediction

- Model 1 (Univariate model) – Wind speed prediction based only on past observation on the same variable.
- Model 2 (Bivariate model) – Wind speed prediction based on wind speed and wind direction historical data
- Model 3 (Bivariate model) – Wind speed prediction based on wind speed and wind gust historical data
- Model 4 (Composite model) – Wind speed prediction based on wind speed, wind direction and wind gust historical data.

Wind Direction prediction

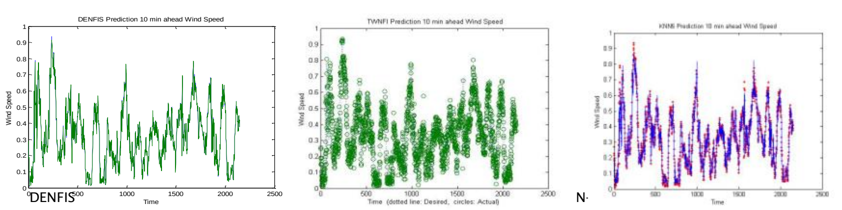
- Wind Direction prediction (univariate) model – Wind direction prediction based only on past observations on the same variable

Wind Gust prediction

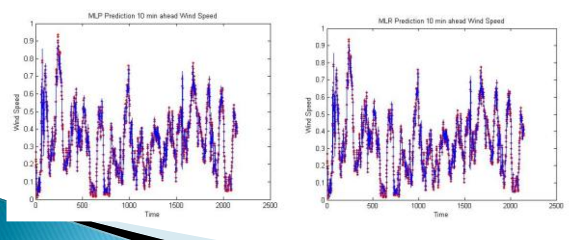
- Wind Gust prediction (univariate) model – Wind Gust prediction based only on past observations on the same variable

Experimental Results

Forecasting horizon: 10 min ahead



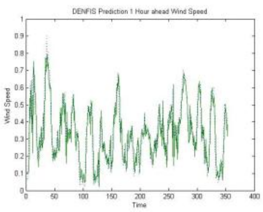
DENFIS N-

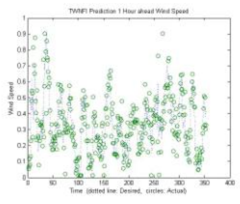


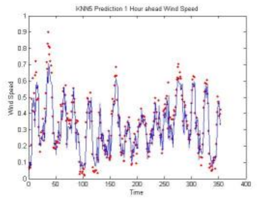
Evaluation 10 min ahead Wind Speed		
Model/Error	RMSE	NDEI
MLR	0.0221	0.1243
MLP	0.0227	0.1279
DENFIS	**0.0218**	**0.1226**
TWNFI	0.0320	0.1780
KNN5	0.0301	0.1694

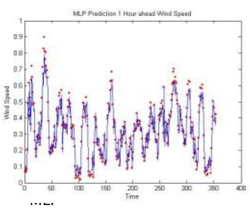
Experimental Results (cont.)

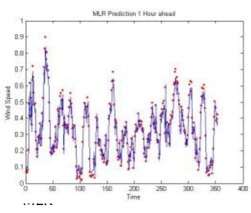
Forecasting horizon: 1 Hour average







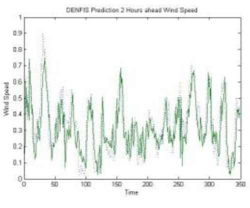


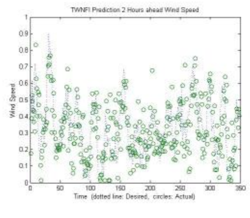


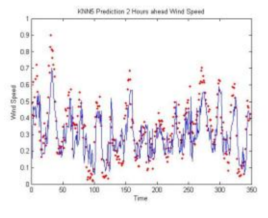
Evaluation 1 Hour ahead Wind Speed		
Model/Error	RMSE	NDEI
MLR	0.0762	0.4391
MLP	0.0763	0.4396
DENFIS	0.0816	0.4701
TWNFI	0.1170	0.6740
KNN5	0.0863	0.4973

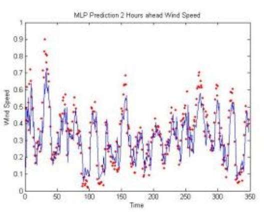
Experimental Results (cont.)

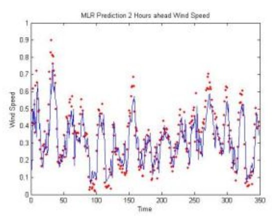
Forecasting horizon: 2 Hours average











Evaluation 2 Hours ahead Wind Speed		
Model/Error	RMSE	NDEI
MLR	0.1067	0.6187
MLP	0.1060	0.6144
DENFIS	0.1622	0.9399
TWNFI	0.2980	1.7290
KNN5	0.1232	0.7141

Summary

- Regarding the training phase DENFIS performs better than any other method, achieving lowest RMSE across all the forecasting horizons
- The accuracy of all the models decreases with the increasing of the forecasting horizon
- The model that achieves optimal error stability and consistency across all the forecasting horizons is DENFIS
- DENFIS satisfies the defined criteria for optimal method selection, because it operates in an online mode, it is able to adapt to new data and changes its structure accordingly.
- DENFIS is regarded as a prominent main forecasting method in the data processing and modelling module of the Integrative Decision Support System Framework.

References

[1] N. Kasabov, "Evolving fuzzy neural networks for online, adaptive knowledge-based learning", IEEE Trans. Syst., Man, Cybern. B, vol. 31, pp.902-918, Dec. 2001

[2] Kasabov, N., Song, Q., "DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and its Application", *IEEE Trans. on Fuzzy Systems*, Vol. 10, No. 2, April, 2002, 144 - 154.

[3] Wu, Yuan-Kang, "A literature review of wind forecasting technology in the world" IEEE Power Tech Conference Proceedings, 2005, 8-15

[4] MATLAB Neural Network Toolbox

THANK YOU FOR YOUR ATTENTION

The Knowledge Engineering and Discovery Research Institute (KEDRI)

Spiking Neural Networks and Quantum-Inspired Particle Swarm Optimisation Methods

Haza Nuzly Abdull Hamed

SPIKING NEURAL NETWORKS AND QUANTUM-INSPIRED PARTICLE SWARM OPTIMISATION METHODS

HAZA NUZLY ABDULL HAMED
Centre for Novel Methods of Computational Intelligence
@ KEDRI, AUT

Primary Supervisor:
PROFESSOR NIKOLA KASABOV
Knowledge Engineering and Discovery Research Institute (KEDRI)
Auckland University Of Technology

Secondary Supervisor:
PROFESSOR SITI MARIYAM SHAMSUDDIN
Soft Computing Research Group (SCRG)
Universiti Teknologi Malaysia

hnuzly@aut.ac.nz

Presentation Outline

Previous work:

1. Evolving Spiking Neural Networks (ESNN)
2. Quantum-inspired PSO (QiPSO)

Proposed methods:

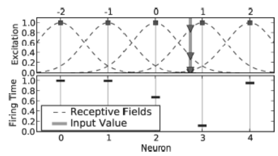
1. Integrated QiPSO-ESNN for simultaneous feature and parameter optimization
2. Dynamic Quantum-inspired Particle Swarm Optimization (DQiPSO)
3. Integrated DQiPSO-ESNN for simultaneous feature and parameter optimization
4. Probabilistic Evolving Spiking Neural Networks (PESNN)
5. Integrated PESNN-DQiPSO for simultaneous connection, feature and parameter optimization
6. Two Layer ESNN for spatiotemporal classification

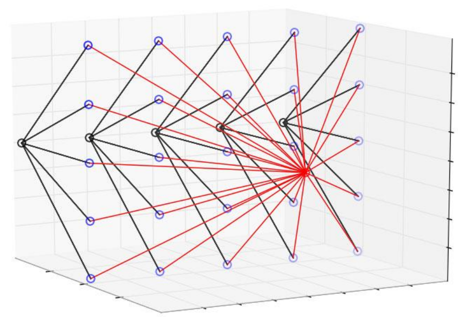
Future work

Previous work - Evolving Spiking Neural Networks (ESNN)

Wysoski, S. G.; Benuskova, L. & Kasabov, N., ICANN 2006

a) Encoding method from real value data to spike time – Population Encoding





b) Neuron model - Thorpe

$$PSP_i = \begin{cases} 0 & if \quad fired \\ \sum w_{ji} * m_i^{order(j)} & else \end{cases}$$

c) Learning method - One-pass learning

Previous work - Quantum-inspired PSO (QiPSO)

Sun, J.; Feng, B. & Xu, W., CEC 2004

The main idea of QiPSO is to use a standard PSO function to update quantum angle θ
$= |\sin(\theta)|^2 + |\cos(\theta)|^2 = 1$

The velocity update formula in standard PSO is modified to get a new quantum angle which is translated to the new probability of the qubit as follows:

$$\Delta\theta_n = w * \Delta\theta_{n_{t-1}} + c_1 * rand() * (\theta_{gbest_n} - \theta_n) + c_2 * rand() * (\theta_{pbest_n} - \theta_n)$$

Then, based on the new θ velocity, the new probability of α and β is calculated using a rotation gate as follows:

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) \\ \sin(\Delta\theta) & \cos(\Delta\theta) \end{bmatrix} \begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix}$$

Equivalent to:

$$\theta = \theta_{t-1} + \Delta\theta$$

where θ is the new quantum angle of the quantum particle position

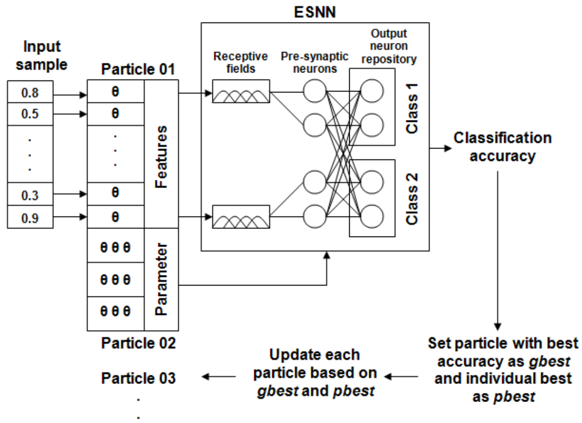
Proposed methods - Integrated QiPSO-ESNN for simultaneous feature and parameter optimization

Haza Nuzly, Nikola Kasabov, Siti Mariyam Shamsuddin, SOCPAR 2009

QiPSO interacts with the ESNN to optimize ESNN parameters and identifies relevant features.

Having fewer significant features could help reduce the processing time and produce good classifications

Inappropriate to manually adjust the parameters especially when dealing with different combinations for different datasets.

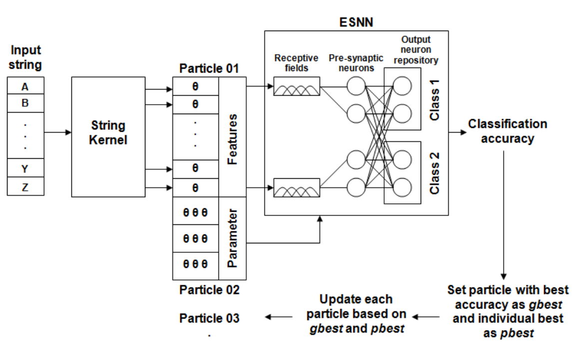


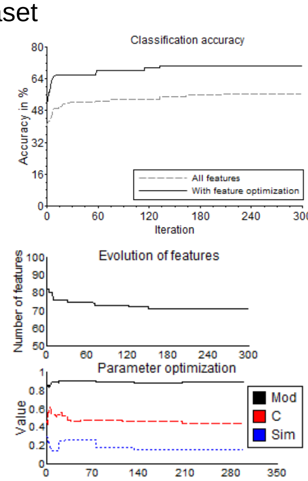
hnuzly@aut.ac.nz

Proposed methods - Integrated QiPSO-ESNN for simultaneous feature and parameter optimization

Haza Nuzly, Nikola Kasabov, Zbynek Michlovský , Siti Mariyam Shamsuddin, ICONIP 2009

In addition to synthetic dataset, the proposed method has been applied to the string classification using Reuters 21578 dataset





hnuzly@aut.ac.nz

Proposed methods - Dynamic Quantum-inspired Particle Swarm Optimization (DQiPSO)

Haza Nuzly, Nikola Kasabov, Siti Mariyam Shamsuddin, ICCSM 2010

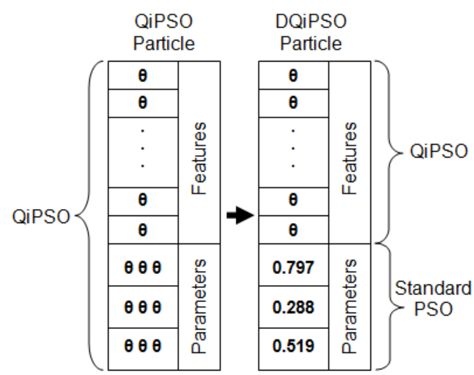
Hybrid Particle Structure

To overcome possibility of missing the optimal ESNN parameter value when using only binary QiPSO.

Example: two qubits to represent parameter with the interval of 0 and 1

4 possibilities:
[00] = 0.0
[01] = 0.3
[10] = 0.7
[11] = 1.0

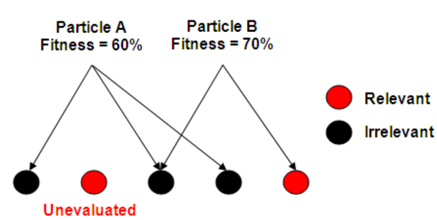
No chance for QiPSO to get the optimal value if the optimal value is 0.8.



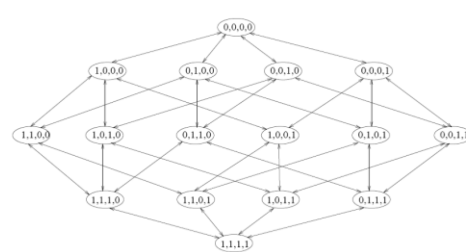
Proposed methods - Dynamic Quantum-inspired Particle Swarm Optimization (DQiPSO)

QiPSO Selection Strategy

QiPSO is based on random selection at the beginning of the process, entire process affected if no relevant feature selected at the beginning

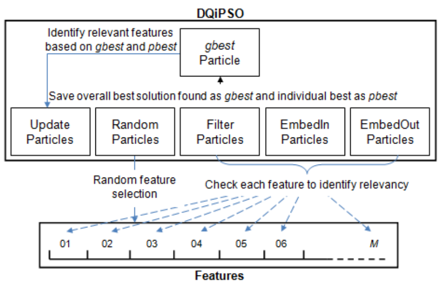


Forward / backward feature selection not suitable for high dimensional problem



Proposed methods - Dynamic Quantum-inspired Particle Swarm Optimization (DQiPSO)

Feature Selection Strategy



Filter, Embed In and Embed Out particles is to identify the relevancy of each feature and reduce the number of candidates until a small subset remains.

For subsequent iterations, features considered relevant will be selected randomly to find the best combination of significant features

Update Particle - Renews itself based on *pbest* and *gbest* information

Random Particle - Randomly generate new sets of features, connections and parameters in every iteration to increase the robustness of the search

Filter Particle - Selects one feature at a time and features with above average fitness will be considered as relevant

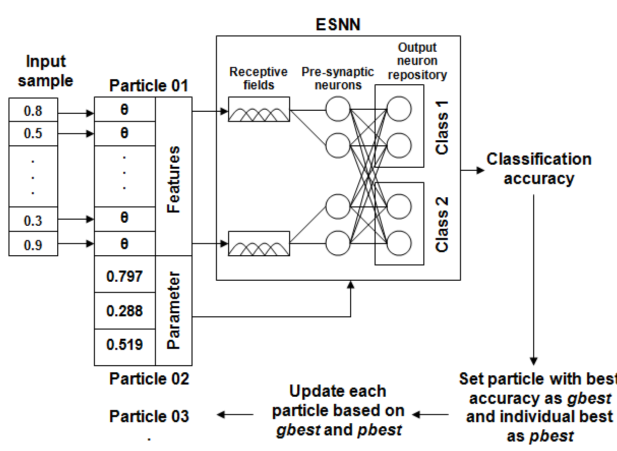
Embed In Particle - Features are added to the network one by one. Considered relevant if the newly added feature improves fitness

Embed Out Particle – Starts all features and gradually removed one by one. Considered relevant if removing a feature causes decrement of the fitness.

hnuzly@aut.ac.nz

Proposed methods – Integrated DQiPSO-ESNN for simultaneous connection, feature and parameter optimization

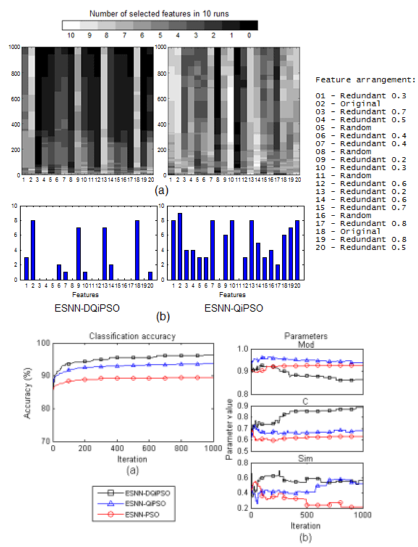
Nikola Kasabov, Haza Nuzly. IJAI 2011
Nikola Kasabov, Haza Nuzly, Siti Mariyam Shamsuddin. Evolutionary Algorithms (BC), 2011



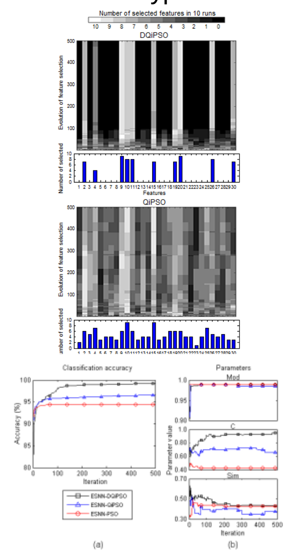
hnuzly@aut.ac.nz

Proposed methods – Integrated DQiPSO-ESNN for simultaneous connection, feature and parameter optimization

Two Spirals



Uniform Hypercube



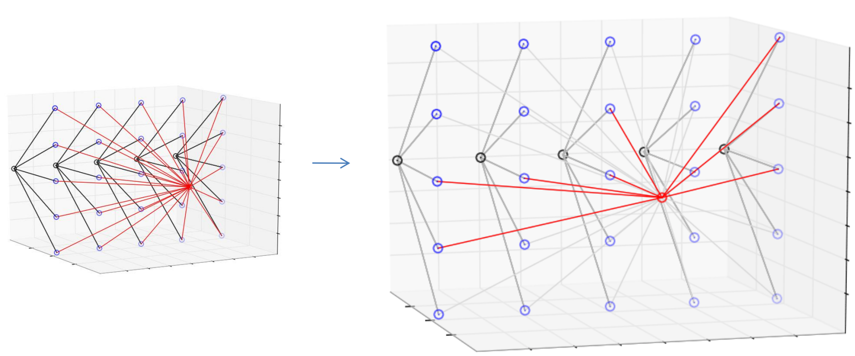
Proposed methods - Probabilistic Evolving Spiking Neural Networks (PESNN)

Haza Nuzly, Nikola Kasabov, Siti Mariyam Shamsuddin, ICONIP 2010

N. Kasabov, “To Spike or Not To Spike: A Probabilistic Spiking Neural Model,” Neural Networks, 2010. 3 probabilities in PNM.

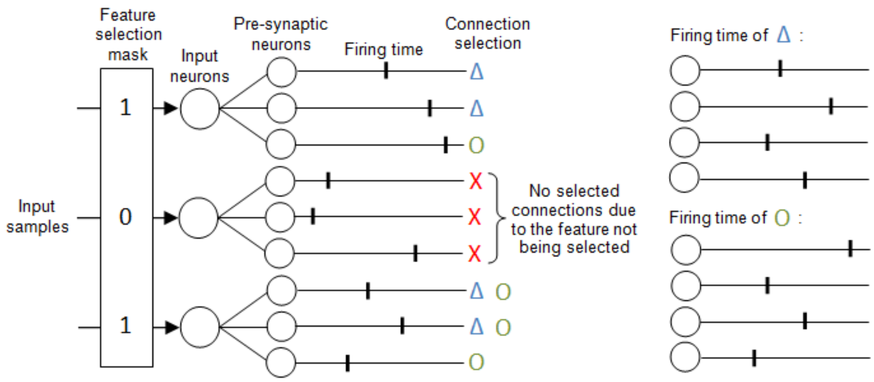
PESNN’s evolving connections = Not all connection exist, not all firing time emit

Let optimizer find the best connections



Proposed methods - Probabilistic Evolving Spiking Neural Networks (PESNN)

By selecting a different connection arrangement, a different firing time will be generated for the pre-synaptic neurons and might give a different output

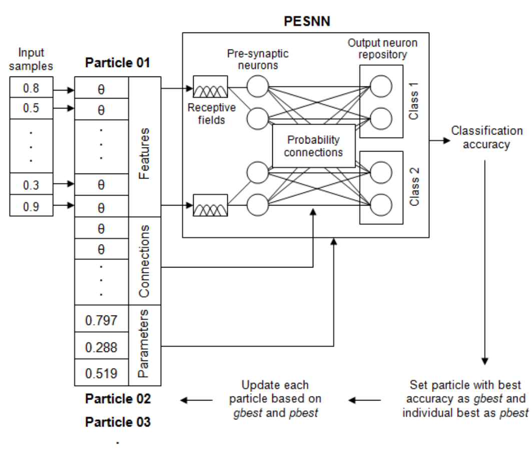


hnuzly@aut.ac.nz

Proposed methods - Integrated PESNN-DQiPSO for simultaneous connection, feature and parameter optimization

Haza Nuzly, Nikola Kasabov, Siti Mariyam Shamsuddin, ICONIP 2010

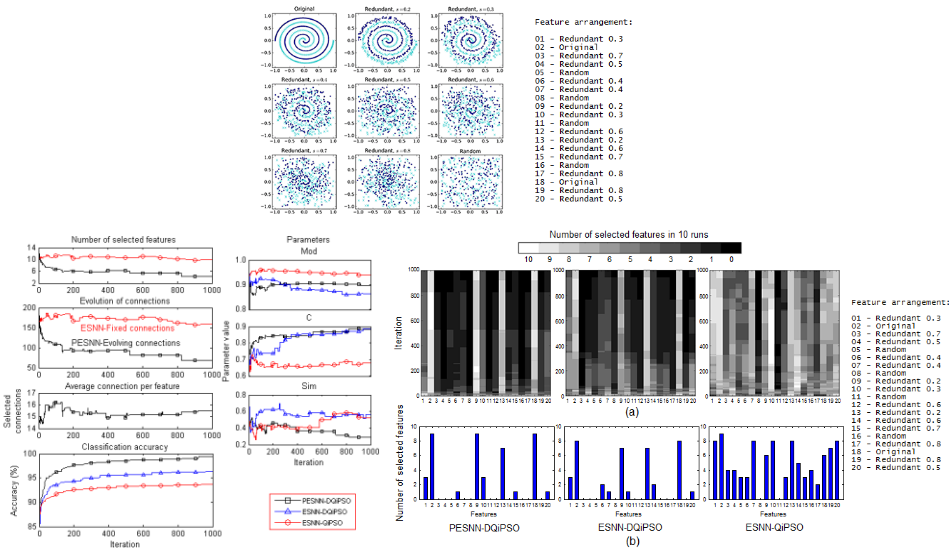
DQiPSO interacts with the PESNN to find best connections, optimize ESNN parameters and identifies relevant features.



hnuzly@aut.ac.nz

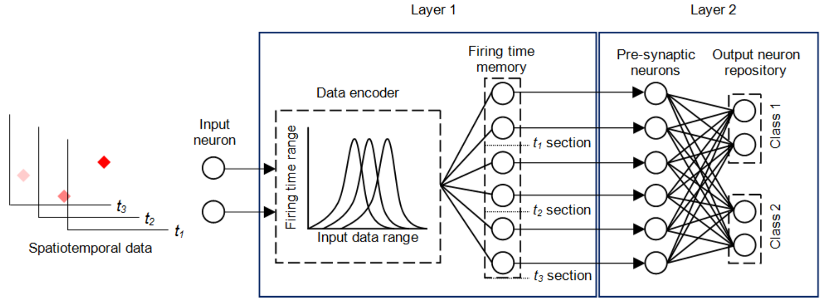
Proposed methods - Integrated PESNN-DQiPSO for simultaneous connection, feature and parameter optimization

Data: 2 Spirals



hnuzly@aut.ac.nz

Proposed methods - Two Layer ESNN for spatiotemporal classification



Definition of a spatiotemporal is a time-evolving spatial object is represented by a set of triplets (o_id, s_i, t_i) where o_id is an object with the identification number, s_i is the location of the o_id at instant t_i .

In spatiotemporal problem, information on spatial is important to represent the event of the object together with the temporal information of timing when the event is happen.

The proposed method is evaluated using LIBRAS sign language movement

hnuzly@aut.ac.nz

Future work

Case study:

1. Integrated PESNN-DQiPSO for species identification – Collaboration with Associate Professor Sue Worner, Lincoln University.

2. Optimize Two Layers ESNN model using DQiPSO

3. Two Layers ESNN for environmental problem – Collaboration with Associate Professor Sue Worner, Lincoln University

hnuzly@aut.ac.nz

THANK YOU

hnuzly@aut.ac.nz

KEDRI's Approach to Spatio-Temporal Data Processing

Dr. Stefan Schliebs

KEDRI's approach to spatio-temporal data processing

Dr. Stefan Schliebs
sschlieb@aut.ac.nz

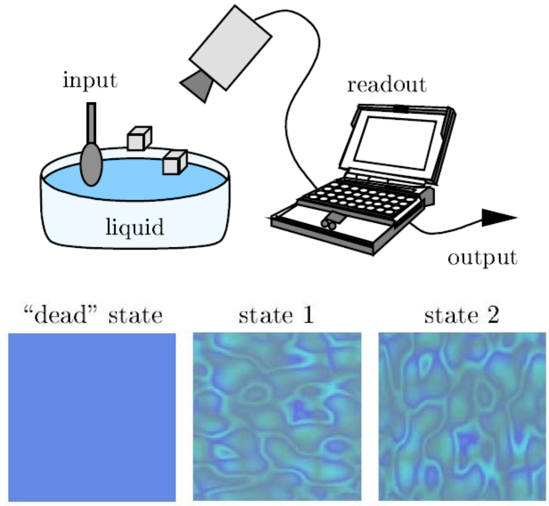
Knowledge Engineering and Discovery Research Institute
www.kedri.info

Overview

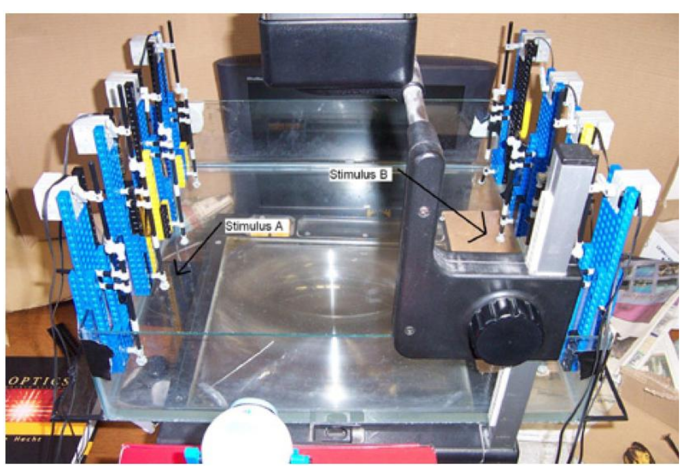
1. Reservoir computing for processing spatio-temporal data
2. Spiking Neural Networks
3. Research projects
4. Software

The Idea of Reservoir Computing

Computation with a cup of coffee?

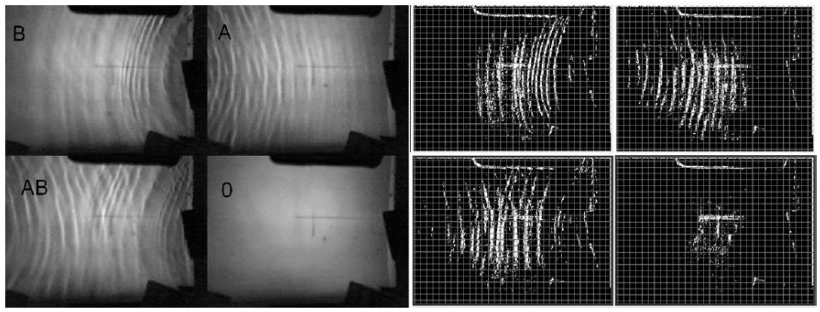


Pattern Recognition in a Bucket!



Fernando, Sojakka, *Pattern Recognition in a bucket*, ECAL'03, LNCS, vol.2801, pp 588-597, 2003

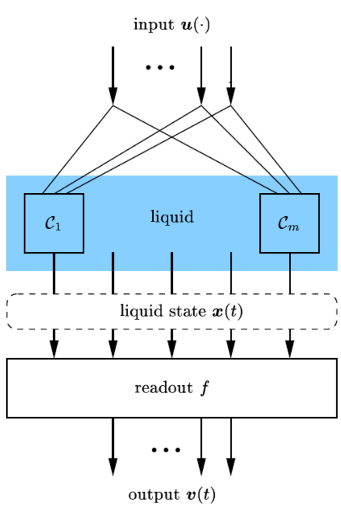
Response of the water bucket on stimulation



Typical pattern for the XOR problem:
top-left: right motor ON → [0,1]
top-right: left motor ON → [1,0]
bottom-left: both motor ON → [1,1]
Bottom-right: right motor OFF → [0,0]

sschlieb@aut.ac.nz

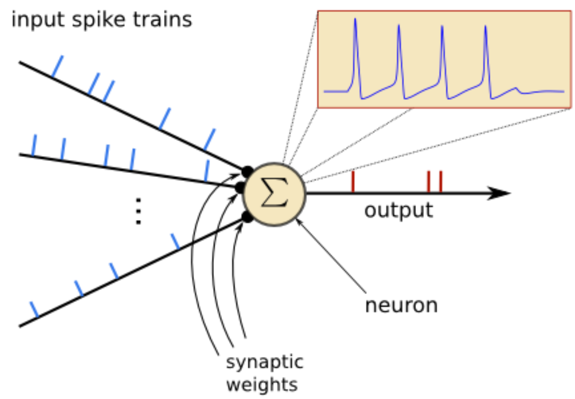
LSM – Liquid State Machine



The "liquid" transforms the input into liquid states *x(t)* which are mapped by a memory-less readout function *f* into a desired target output *v(t)=f(x(t))*.

sschlieb@aut.ac.nz

Spiking Neuronal Models

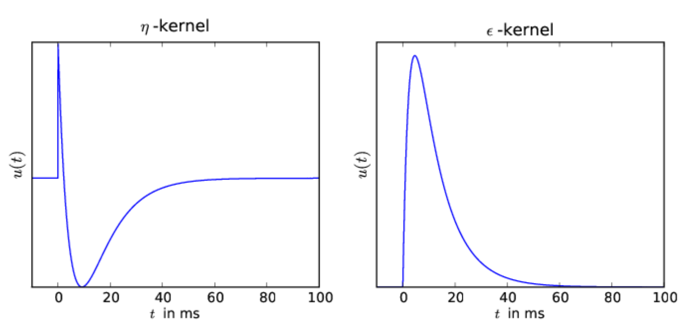


sschlieb@aut.ac.nz

Example: Spike Response Model

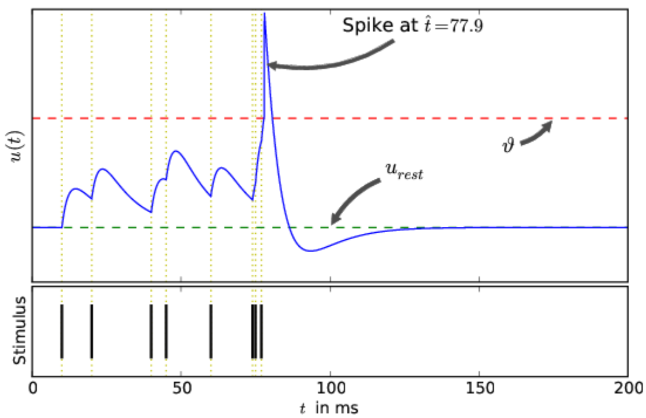
$$u(t) = \eta(t-\hat{t}) + \sum_j w_j \sum_f \varepsilon_j(t - t_j^{(f)})$$

where $t_j^{(f)}$ are the firing times of presynaptic neurons j, while w_j represents the synaptic efficacy between the considered neuron and neuron w_j

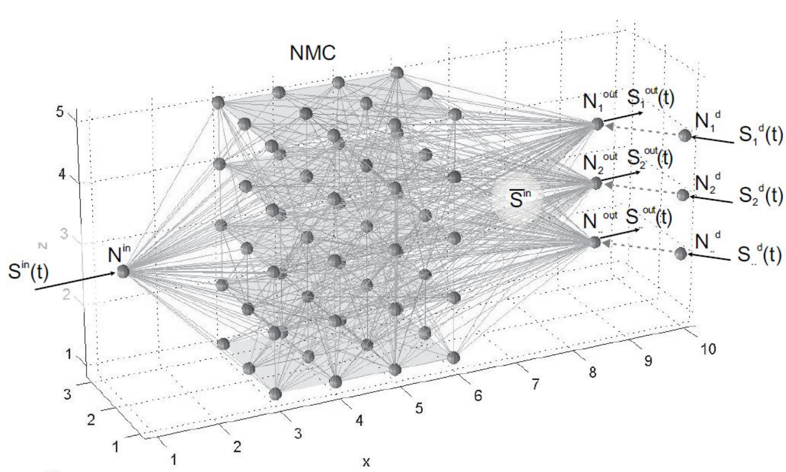


sschlieb@aut.ac.nz

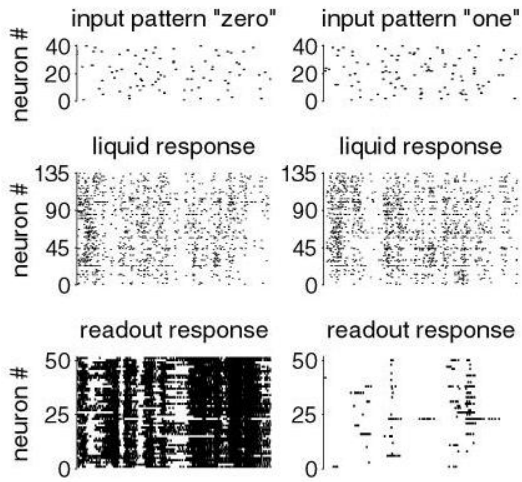
Dynamics of the Spike Response Model



LSM Topography

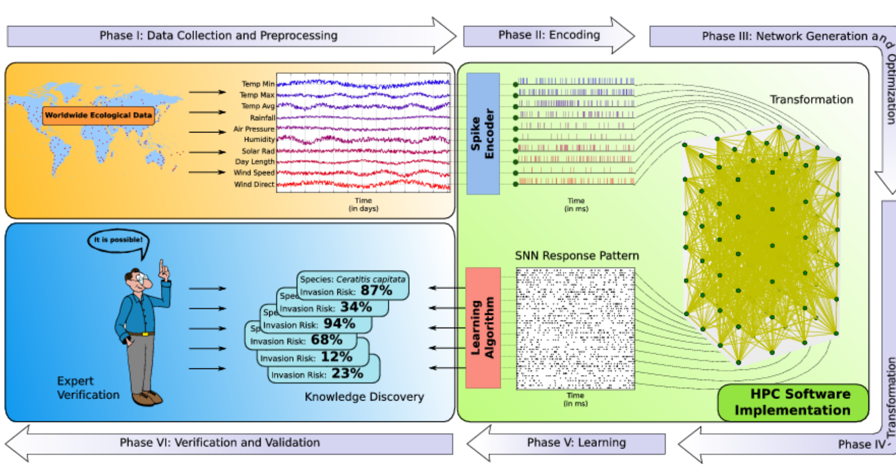


Liquid Response



sschlieb@aut.ac.nz

Applying Advanced Neural Computing to Ecological Problems



sschlieb@aut.ac.nz

Software

Demo

sschlieb@aut.ac.nz

Integrated Optimization Method for Personalized Modelling

Dr. Yingjie (Raphael) Hu

Integrated Optimisation Method for Personalised Modelling

Dr. Yingjie (Raphael) Hu

KEDRI, AUT

February 7, 2011

Table of contents

Background

1. Global Modelling
   Most contemporary decision support systems use global models for prediction problems. A global model is derived from all available data for the target and then applied to any new patient anywhere at anytime.
2. The Issue
   Prediction and treatment based on global models are only effective for some patients ($\approx 70\%$)

Outline
Introduction
Research Objectives
Methodology
Case Study
Conclusion

Background

1. Global Modelling
   Most contemporary decision support systems use global models for prediction problems. A global model is derived from all available data for the target and then applied to any new patient anywhere at anytime.
2. The Issue
   Prediction and treatment based on global models are only effective for some patients ($\approx 70\%$)
3. Personalised Modelling
   The rationale behind personalised modeling paradigm is: since each person is different, the most effective treatment could be only based on the detailed analysis for this particular patient.

Dr. Yingjie (Raphael) Hu Integrated Optimisation Method for Personalised Modelling

Background

1. Global Modelling
   Most contemporary decision support systems use global models for prediction problems. A global model is derived from all available data for the target and then applied to any new patient anywhere at anytime.
2. The Issue
   Prediction and treatment based on global models are only effective for some patients ($\approx 70\%$)
3. Personalised Modelling
   The rationale behind personalised modeling paradigm is: since each person is different, the most effective treatment could be only based on the detailed analysis for this particular patient.

Research Objective

1. To address the issues in personalised modeling for data analysis, such as imbalanced data class distribution, feature selection, classification models, parameter optimisation, error measuring, etc;
2. To create an accurate personalised computational model using information for an individual and the available information for other individuals that is related to the same problem;

Research Objective

1. To address the issues in personalised modeling for data analysis, such as imbalanced data class distribution, feature selection, classification models, parameter optimisation, error measuring, etc;
2. To create an accurate personalised computational model using information for an individual and the available information for other individuals that is related to the same problem;
3. To develop new algorithms and methods for personalised modelling;

Outline
Introduction
Research Objectives
Methodology
Case Study
Conclusion

Research Objective

1. To address the issues in personalised modeling for data analysis, such as imbalanced data class distribution, feature selection, classification models, parameter optimisation, error measuring, etc;
2. To create an accurate personalised computational model using information for an individual and the available information for other individuals that is related to the same problem;
3. To develop new algorithms and methods for personalised modelling;
4. To apply the above proposed algorithms and methods on the data from different sources, such as gene expression data, protein data, SNPs (single-nucleotide polymorphism) data, clinical data, etc;

Research Objective

1. To address the issues in personalised modeling for data analysis, such as imbalanced data class distribution, feature selection, classification models, parameter optimisation, error measuring, etc;
2. To create an accurate personalised computational model using information for an individual and the available information for other individuals that is related to the same problem;
3. To develop new algorithms and methods for personalised modelling;
4. To apply the above proposed algorithms and methods on the data from different sources, such as gene expression data, protein data, SNPs (single-nucleotide polymorphism) data, clinical data, etc;

Outline
Introduction
Research Objectives
Methodology
Case Study
Conclusion

Research Objective

1. To address the issues in personalised modeling for data analysis, such as imbalanced data class distribution, feature selection, classification models, parameter optimisation, error measuring, etc;
2. To create an accurate personalised computational model using information for an individual and the available information for other individuals that is related to the same problem;
3. To develop new algorithms and methods for personalised modelling;
4. To apply the above proposed algorithms and methods on the data from different sources, such as gene expression data, protein data, SNPs (single-nucleotide polymorphism) data, clinical data, etc;

Dr. Yingjie (Raphael) Hu Integrated Optimisation Method for Personalised Modelling

The IMPM Method

- An Integrated Method for Personalised Modelling (IMPM) Utilising Features, Model Parameters and Neighbourhood Optimisation;

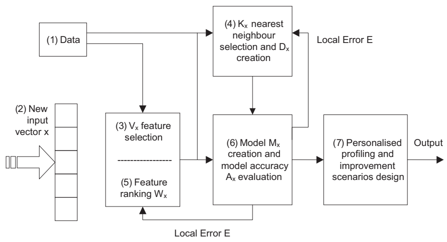


Figure: A functional block diagram of the proposed IMPM

IMPM for Colon Cancer Diagnosis and Profiling - 1

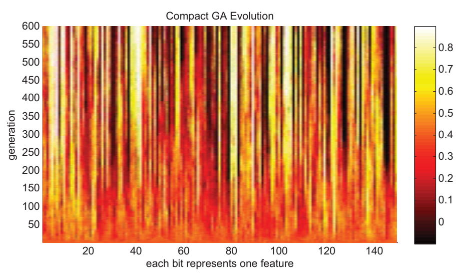


Figure: The evolution of feature (variable) selection for sample#32 from the Colon cancer data (600 generations of GA optimisation; the lighter the colour, the higher the probability of the feature to be selected; each feature is represented as one bit on the horizontal axis; at the beginning all features are assigned equal probability to selected as 0.5)

IMPM for Colon Cancer Diagnosis and Profiling - 2

Table: The 20 most frequently selected genes (potential marker genes) using the proposed IMPM across all colon cancer gene data samples

Index of Gene	GenBank Accession Number	Description of the Gene (from GenBank)
G377	Z50753	H.sapiens mRNA for GCAP-II/uroguanylin precursor
G1058	M80815	H.sapiens a-L-fucosidase gene, exon 7 and 8, and complete cds.
G1423	J02854	Myosin regulatory light chain 2, smooth muscle ISOFORM (HUMAN)
G66	T71025	Human (HUMAN)
G493	R87126	Myosin heavy chain, nonuscle (Gallus gallus)
G1042	R36977	P03001 Transcription factor IIIA
G1772	H08393	COLLAGEN ALPHA 2(XI) CHAIN (Homo sapiens)
G765	M76378	Human cysteine-rich protein (CRP) gene, exons 5 and 6.
G399	U30825	Human splicing factor SRp30c mRNA, complete cds.
G1325	T47377	S-100P PROTEIN (HUMAN).
G1870	H55916	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE, MITOCHONDRIAL PRECURSOR (HUMAN)
G245	M76378	Human cysteine-rich protein (CRP) gene, exons 5 and 6.
G286	H64489	Leukocyte Antigen CD37 (Homo sapiens)
G419	R44418	Nuclear protein (Epstein-barr virus)
G1060	U09564	Human serine kinase mRNA, complete cds.
...	...	...

IMPM for Colon Cancer Diagnosis and Profiling - 3

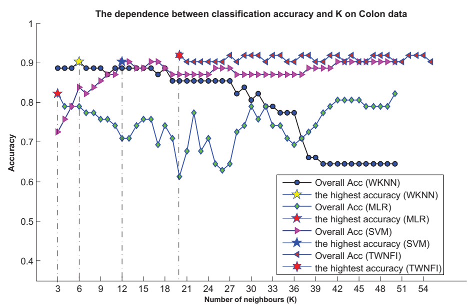


Figure: A comparison of classification results obtained by 4 classification algorithms using 20 potential maker genes, where x axis represents the size of neighbourhood and y axis is the average classification accuracy across all samples. The best accuracy is obtained with the use of the TWNFI classification algorithm (91.90%)

Conclusion

- Compared to global or local modelling, the proposed IMPM has a major advantage:
  the modelling process starts with all relevant variables available for a person, rather than with a fixed set of variables required by a global model, which may not be necessarily representative in terms of best prognosis for this person;
- The proposed IMPM leads to a better prognostic accuracy and a computed personalised profile;

Outline
Introduction
Research Objectives
Methodology
Case Study
Conclusion

Conclusion

- Compared to global or local modelling, the proposed IMPM has a major advantage:
  the modelling process starts with all relevant variables available for a person, rather than with a fixed set of variables required by a global model, which may not be necessarily representative in terms of best prognosis for this person;
- The proposed IMPM leads to a better prognostic accuracy and a computed personalised profile;
- With global optimisation using IMPM, a small set of variables (potential markers) can be identified from the selected variable set across the whole population;

Conclusion

- Compared to global or local modelling, the proposed IMPM has a major advantage:
  the modelling process starts with all relevant variables available for a person, rather than with a fixed set of variables required by a global model, which may not be necessarily representative in terms of best prognosis for this person;
- The proposed IMPM leads to a better prognostic accuracy and a computed personalised profile;
- With global optimisation using IMPM, a small set of variables (potential markers) can be identified from the selected variable set across the whole population;
- The proposed algorithms and models of IMPM are generic which can be potentially incorporated into a variety of applications for data analysis and knowledge discovery with certain constraints, such as financial risk analysis, time series data prediction, etc;

Outline
Introduction
Research Objectives
Methodology
Case Study
Conclusion

Conclusion

- Compared to global or local modelling, the proposed IMPM has a major advantage:
  the modelling process starts with all relevant variables available for a person, rather than with a fixed set of variables required by a global model, which may not be necessarily representative in terms of best prognosis for this person;
- The proposed IMPM leads to a better prognostic accuracy and a computed personalised profile;
- With global optimisation using IMPM, a small set of variables (potential markers) can be identified from the selected variable set across the whole population;
- The proposed algorithms and models of IMPM are generic which can be potentially incorporated into a variety of applications for data analysis and knowledge discovery with certain constraints, such as financial risk analysis, time series data prediction, etc;
- We hope that this study will motivate the the applications of personalised modelling research in different research areas.

Dr. Yingjie (Raphael) Hu Integrated Optimisation Method for Personalised Modelling

Conclusion

- Compared to global or local modelling, the proposed IMPM has a major advantage:
  the modelling process starts with all relevant variables available for a person, rather than with a fixed set of variables required by a global model, which may not be necessarily representative in terms of best prognosis for this person;
- The proposed IMPM leads to a better prognostic accuracy and a computed personalised profile;
- With global optimisation using IMPM, a small set of variables (potential markers) can be identified from the selected variable set across the whole population;
- The proposed algorithms and models of IMPM are generic which can be potentially incorporated into a variety of applications for data analysis and knowledge discovery with certain constraints, such as financial risk analysis, time series data prediction, etc;
- We hope that this study will motivate the the applications of personalised modelling research in different research areas.

Outline
Introduction
Research Objectives
Methodology
Case Study
Conclusion

Thank You.

Dr. Yingjie (Raphael) Hu Integrated Optimisation Method for Personalised Modelling

East China - West China - New Zealand Tripartite Workshop

Publishing Assistant: Kshitij Dhoble

www.kedri.info