Conference note template

Jung Xue

2020-11-28

Contents

Conference information

XXXX Conference:

- **Time:** 8:55 Tuesday 24/11/2020 Wednesday 25/11/2020
- **Venue:** MLT2/303-102 Map
- **Registartion:** Yes
- **Hosted by:** NZSA
- **Organiser:** Organiser Email
- **Conference Schedule** Link Here
- **Extra** AGM meeting at 12:30 Link here

Keynote Speakers:

Speaker	Topic………………………Email	Website
Chris Wild	Education democratizing data and software Targeting the intersection	
Felipa Zabala	A framework to evaluate imputation strategies at Stats NZ	
Susmita Das	A machine learning model to identify private dwellings from admin data	
Simon Urbanek	Interactive Visualisation using RCloud	

Speaker	Topic	Email	Website
Jason Wen	Accessing evidence of firing pin impression by using machine learning	jwen246@ aucklanduni.ac. nz	
Richard Penny	Modelling for COVID in Official Economic Time Series		
Maree Luckman	A lifetime of data - Biometrics Technician to Senior Applied Statistician		
Andrew Balemi	There and back again: A statisticians journey into the ‘real world’ and back to academia		
Agnes Yongshi Deng	Designed experiments for tuning hyperparameters in machine learning algorithms	yongshi.deng@ auckland.ac.nz	
Alistair Ramsden	Testing the confidentiality of synthetic data for the Stats NZ Integrated Data Infrastructure (IDI) Population Explorer dataset		
Rory Ellis	Using Bayesian Growth Models to Predict Grape Yield		

Speaker	Topic........................	Email	Website
Martin Hazelton	The Future of Statistics at New Zealand Universities		
Wilma Molano	HLFS mode of collection: A journey due to COVID-19		
Shanika Wickramasuriya	Non-negative forecast reconciliation for forecasting hierarchical time series	s.wickramasuriya@auckland.ac.nz	
Claudia Rivera-Rodriguez	Optimal sampling allocation for outcome dependent designs in cluster-correlated data settings		
Martin Upsdell	Estimating the time lag between predator abundance and prey abundance		
Richard Arnold	Statistics of Ambiguous Rotations		
Len Cook	Missing in action - a statistical window on prisons		
Peter Mullins	War Stories	len_cook@xtra.co.nz	https://www.wgtn.ac.nz/igps/about-us/staff/senior-associates/mr-len-cook

Speaker	Topic…………………….Email		Website
Thomas Lumley	Influence functions, and why you should care		
Beatrix Jones	Dimension reduction for imbedding high dimensional measurements into Bayesian Networks		
Alasdair Noble	A Bayesian approach to modelling of Phosphorus inputs to rivers from diffuse and point sources		
Andrew Sporle	Beyond the Integrated Data Infrastructure - building a strategic data resource for Aotearoa		
Azam Asanjarani	Decision Making for Partially Observable Markov Processes		

interesting people I have meet/noticed

People	Field/Job	Contact	Facts
Anna?	PhD @ Otago		Likes Pythagoras and median theory, works on musclefiber study

Note: All information disclosed within this conference e-note are intented for personal use.

Chris Wild | Education, democratizing data, and software: Targeting the intersection

0.1 **Democratising data: *EMPOWERING THE MANY**

Definition

Access, Capability can not do without another

enable decision makers

IDI world leading data system

Linkable

Problems

increasingly complex, less accessable

use IDI to inform, promated by government

technical barrier, not user friendly

offocial stats increasingly not for the people,

administrative/business data

missing/incomplete data

What we can do about it

data informed decision to population

once data is lost, you cannot get it back indigenous data sovereignty

make sure data get used, in a positive way

ITI Information access and governiance Translator Imfrastrustur eto make it happen

reduce technical barriers

Paper: Indigenous data sovernity and policy

0.2 Wild

Education vs software my thought: booking ticket example, use to have to go to agernts,now just app do it fast and good enough, better than slow and perfect

enable people who coding is not a serius option enable practical cababilities

gard school stuff, exposure in high school grad school sruff is something most people never learn about

show and tell (graph an d summary)

high level instructions default answers context aware choices

beginner friendly

Alot of advantage for coding

Flexibilty, reproducibility, long run time friendly, history track, able to deal with large and complex data,

my thought:AI assisted Rstudio? Telling you what to do next, what options do you have, where seem to have typo/bug, auto alignment

Felibel Zabala | A framework to evaluate imputation strategies at Stats NZ

0.3 Subsection

chalenge of big data

ereous and missing data Greater problem duing COVID

treat and calculate using existing data

Desired properties

Predictive accuracy ranking accuracy distribution accuracy estimation accuracy imputation plausibility: impyationvalues that are plausible

clean dataset to assess imputation

response mechanism

pearsons correlation, good impytation should have R2 close to 1

Household Evalaution Survey 2015 first used

income, age sex etc+ demographics

imputation method Nearest neighbour mean impouation with error term

Key variable: income

Larger weight for more significant variables,

standarised codes for evaluation compuations of estimation of bias variance and mean square error

felipa.zabala@stats.govt.nz

Susmita Das | A machine learning model to identify private dwellings from admin data

0.4 Back ground

census immuerate resident dwellimg and collect attributes

can we not use census, instead use administravive data

2018 had lower thane xpected response rate

can we move into fullt administarvive census

current collection method, address list, info to make sure address is up to date,

obj of study: model to predict private dwelling from administative data

prison, businesss, not private dwelling

why predict dwelling

address will not tell you whether this is a dwelling or not

address have insights

assumptions: every address is unique dwelling work as privcate residence address linked in admin data is current

work with anomaised address for security reason

need to dicide for threshld

Future work

more data source Assessment of assumptons selection of trainning and validation dataset

Simon Urbanek | Interactive Visualisation using RCloud

Fantastic beard

0.5 Subsection

data and analytics in the cloud webabsed sharing and collaboration

Support web graphics PROBLEM WITH NO INTERNET

Jason Wen | Accessing evidence of firing pin impression by using machine learning

0.6 Subsection

Image processing to improvedata quality

ie zoom in area of interest

histogram equalization (improve contrast)

noise reduction using filters

mage enhancing algorithms

summarise feature into 1 d

histogram of orientated gradient (HOG)

gain 2 d image gradient turninto a histogram

hetrogenity, image from same sourece more similar

local HOg feature comparison

Richard Penny | Modelling for COVID in Official Economic Time Series

0.7 NZSTATS

5000 time series per quarter must be robust and automated

mainly seasonally adjusted trend and estimated

TS expert adjust variables, result for users

time series must be consistent, people hated change

need to do things right the forst time

COVID affect different places at different time

RegARIMA,

add a covid variable B_iX_it

possible covid effects outlier, level shift, ramp, step functoion

not enough data, noise, etc made it hard to identify the trend

red flages, covid values made so much impact seasonality started to signii=ficantly shift

visitor arrivals goes to 0

we cannot assume any model for COVID priod, and may have profound long term effect for future arrivals

what can we expect

seasonal breaks

merging data?

North and south hemisphere seasonality??

Maree Luckman | A lifetime of data - Biometrics Technician to Senior Applied Statistician

0.8 Facing her challemges

What is your motivator in life?

Share the enthusiasm with your collegues

Technology changes ways scientists work

quality of data

Find the white space, where consumer needs a product byt the product does not exist

no such thing as short question

okay to say no, okay to say that I will get back to you on that

define the problem in the way that you can assess

be open am d honest,open abloyt llearning

trainning session

hwo do you feel about it

excellent question and interrorigsting skill

Active listenin skills

look like that you are doing some serious work, buzz words,

communicatonis to gather informationa dnsolve problems

Natural curiocity is important

trust in consulatnat cient relationshipsse why?

collaborative cooprtation

dat scientist vs statistician

Andrew Balemi | There and back again: A statisticians journey into the 'real world' and back to academia

0.9 Subsection

Real world intimidating

mistakes teach you the most

yes or no, remember ads

no advertising, exponential decay

wasn't exact, but did a good job

you dont have to make things complicated, somrtime easy solutions works too

you have already been taught of the solution

text book are the worst place to get inspirations from

is this real or bullshit, be critical

confirm your results with your clients

what you do know is a good place to start, and add complexity as you need it

effect of promotion?

Andrew overfitting, walking down to car,

THEORY INFORMS APPLICATION

good and effective eway to be lazy

listening

get out of your comfort zone

Agnes Yongshi Deng | Designed experiments for tuning hyperparameters in machine learning algorithms

0.10 Subsection

Alistair Ramsden | Testing the confidentiality of synthetic data for the Stats NZ Integrated Data Infrastructure (IDI) Population Explorer dataset

0.11 Subsection

Rory Ellis | Using Bayesian Growth Models to Predict Grape Yield

Prediction of grape yield base on seasonal factors and industry practiss

Grape vie 2 year cycle

assume no neg

tak log response

double sigmoidal model

impact of incorporating historical data

less volitalitywhen incorporating historical data

Vague prior , informed prior

Bayesian model is sensitive to prior assumptions

tradeo off in early and late year predictions

idea of the bucket problem

loses of v ariation between indic=vidual masses,

Maybe check out his papers

Martin Hazelton | The Future of Statistics at New Zealand Universities

0.12 Subsection

Statistics is going a chaanging time Data scientist error

analytic, data science, data mining, machine learning etc

Budgetting of university and government

some ideas that statisticians value not necessary what some other believe

Less practical, loses competitive edge

rebranding carries a long time risk

buzz words

research assessment exercises (in UK)

PBRF funding (in NZ)

increase of staff in top universities, strengthern by ranking

but overall number of stats department and staff drops

diaspear in the pit of death /9less tha 6 staff)

University of Auckland is doing well

Hardly a panic situation,

Challenges

low research profile poor majoring numbers honours trradtiotion way to get to PHD,

Solutions?

more people from non-traditional background data science is a opportunity

tragedy at waikato

had math stats CS, but did not build on that but became competitive bertween CS and stats

Wilma Molano | HLFS mode of collection: A journey due to COVID-19

0.13 Subsection

STATS2 NZ stopped face to face interview 20 March 2020

June quarter

1/8 F2F 7/8 Call center

possible source of bias due to COVID

i.e. unemployment rate effect of data collection mode

unemployment rate filed higher than call center. field interview tend to pick up more younger people, more Maori

letters + reminders increased response rate

look at thing sin advance monitoring during and after

Shanika Wickramasuriya | Non-negative forecast reconciliation for forecasting hierarchical time series

0.14 Subsection

0.15 Further reading

https://robjhyndman.com/publications/nnmint/

Check out her slide

very hard

Claudia Rivera-Rodriguez | Optimal sampling allocation for outcome dependent designs in cluster-correlated data settings

0.16 intro

Allocation => given N, how many do we sample?

we are interested in regression

weighted genealised estimating equation

remind your self about minimizing

My thought: can tables be more concise and clear compare to multipple similar plots?(ie heat map)

Martin Upsdell | Estimating the time lag between predator abundance and prey abundance

0.17 Subsection

irish wasp (predator)

clover root weevil

predator and prey curve should match, after shifting time lag and stanadise the numbers

Richard Arnold | Statistics of Ambiguous Rotations

0.18 Subsection

Len Cook | Missing in action - a statistical window on prisons

0.19 Subsection

Peter Mullins | War Stories

0.20 Subsection

Thomas Lumley | Influence functions, and why you should care

0.21 Subsection

Beatrix Jones | Dimension reduction for imbedding high dimensional measurements into Bayesian Networks

0.22 Subsection

Alasdair Noble| A Bayesian approach to modelling of Phosphorus inputs to rivers from diffuse and point sources

0.23 Subsection

Andrew Sporle | Beyond the Integrated Data Infrastructure - building a strategic data resource for Aotearoa

0.24 Subsection

Azam Asanjarani | Decision Making for Partially Observable Markov Processes

0.25 Subsection

Concluding Remarks

What did you learnt by the end of this session/course?

Take home message?

Add 3 questions to ponder.

How to use RBookDown

Firstly, you must read the RBookDown Bible by YiHui Xie

In essence, you write in a mixture of markdown (For basics), html (to extend on markdown) and latex language (mostly for equations) to create a simple Note.

You can customise your style and theme through your own CSS.

RMarkdown are mostly preferably used to knit e-books(HTML), use TexStudio if you want a proper printable PDF, Latex will be easier.

Here are some useful tips to get started

1: To add a chapter, just open a R file and save as `.RMD`. Use number 0 to 99 with a hyphen `-` to order the RMD files and maybe add a Chapter name so it is easier to select from `Files` window at bottom right of the R Studio.

2: Code chunks can generate graphical outputs, To insert pictures just use `include_graphics` instead of `\includegraphics{}` or `![]()`. Width can be customised.

```
knitr::include_graphics(rep('images/knit-logo.png', 3))
```

3: Use 1 grave accent ‘ to include the inline code, use 3 grave accent to include a chunk of code.

4: use `{-}` to stop automatic chapter names

5: Often you have tables, you can copy the table to a excel file and convert table to markdown tables, using Online Websites