S1.

Hi everyone and good morning. Firstly let me thank everyone for making the time today for my CA1 presentation and before I begin I would like to introduce myself.

I am Mark Caple and after many years of working in the IT industry

from developer through to architect I was lucky enough in 2019 to be accepted by the School of Computer Science at UTS to undertake a Masters by Research. The research is in collaboration with the industry partner Workforce Health Assessors. It is supervised by Associate Professor Farookh Hussain and co-supervised by Associate Professor Fan Dong.

So let’s jump in

<click>

S2.

Firstly let me describe the general outline of everything we will cover today:

<click>

To start I will cover some background information on my topic

<click>

A literature review will describe how I discovered gaps in the current literature

<click>

We will then describe the those gaps

<click>

From the gaps we will define our research questions

<click>

Our objectives will follow this

<click>

From here we will talk more deeply about how we will attempt to answer these questions in our research methodolgy

<click>

We will in the final part of the presentation talk about any research significance and

<click>

Finally we will discuss the research plan and at the end of the presentation a short demo will elaborate on our progress to date

<click>

**S4.**

**The humble Questionnaire although commonly used today has a history of less than 200 years. It’s origins have been attributed to the Statistical Society of London in 1838 whose main goal was ”procuring, arranging and publishing facts to illustrate the condition and prospects of society”**

**They wanted to improve methods of gathering what they called statistics for life and to recommend these objects of inquiry to which attention maybe profitably directed.**

**Since those beginnings we have further attempted to categorise the questionnaire and Marshall summarised the questionnaire into two broad categories**

**<click>**

**1.** **Open-ended – where the question allows the participant to enter whatever they feel is**  **appropriate with no restrictions being**  **introduced.**

***What do you like to do on the weekend?***

And

<click>

**2.** **Closed-ended – where the question would have one correct answer or a limited number of**  **options.**

***Select the things you do on the weekend?*  1. Read, 2. Housework, 3. Eat out**

I hope that you are able to deduce from these two definitions **that open-ended questions would tend to promote long responses whereas closed-ended short responses.**

**A great deal of time and effort in recent times has been channelled into mining trends, sentiment and general information from free text using a variety of NLP techniques. Our research is focused on the second type of closed-ended questionnaire which is of particular interest to our industry partner.**

**<click>**

**S5.**

**Marshal goes on to say that the answers to closed-ended questions maybe** further split into five different types

1. category - represents a set of mutually exclusive categories (e.g male, female)

2. list - multiple category choice is possible as the answers are not exclusive,

eg ”what services have you used from your GP in the last year?”.

3. quantity/numeric - such as ”how many times have you broken your leg?”

4. ranking/scale - such as ”how would you rate your doctor [1-7]”

5. linguistic ranking/scale - such as ”would you describe yourself as: very tall,

tall,short,very short?”

The important thing to remember here is that a closed-ended questionnaire maybe made up of all or only a few of these types.

<click>

S7.

Our university industry partner Workforce health assessors has a core service that is pre-employment assessments.

<click>

Currently they offer third party organisations an efficient means to bring candidates onboard which can involve interview(s), medical questionnaires and medical assessments.

Their current system allows for specific role based questionnaires to be created for an individual client. The process of candidate selection would typically start with an interview which could also be accompanied with a medical questionnaire and lastly a medical assessment.

The reason for this order is that only probable candidates would be required to go through a potentially expensive medical assessment.

<click>

Through experience however they have realised that sometimes a candidate that appears perfect for a role fails at the last hurdle of a medical assessment. It is this late assessment failure that brings rise to the core problem of this research. How can a potential candidate be assessed on some medical criteria without involving an actual medical assessment?

By focusing on the medical questionnaire it is hoped that a sufficient number of medical assessments will not be required leading to the removal of a number of negative outcomes. These range from having to select another possible candidate who has since found alternate work through to the possibility of having to start the whole recruitment campaign again due to the lack of alternate candidates.

<click>

S9.

A potential candidate must first discover a role or job that they are interested in but this job maybe named differently by different people. For instance a bus driver could also be referred to as a bus operator or omnibus driver and so a standard convention is required. This convention within Australia and New Zealand is the Australian and New Zealand Standard Classification of Occupations (ANZSCO). This ANZSCO standard will form the ”bridge” between the description of a role given by a candidate and the definition of the role within the system.

<pause>

<click>

The system then connects each ANZSCO code to one or more risk profile templates.

A risk profile forms an association between a particular job requirement such as ”lifting heavy weight from floor to waist” or ”sitting for extended periods” through to the body parts that are affected due to that requirement, such as back, arm, leg or shoulder.

<click>

S10.

From these body parts a number of common injuries are linked and from these injuries ultimately a series of questions are added to the questionnaire for that role.

In this figure shown we see two job requirements 1 & 2 that follow the classic association just described whereas job requirement 3 links directly to multiple question groups. An example of such a requirement would be a bus driver where a particular licence is a requirement which obviously has no reliance on a body part.

<click>

S11.

The following shows a specific job requirement wherein a requirement of lifting heavy weight from floor to waist affects the hip which in turn is susceptible to hip flexor strain and this particular injury is connected to a hip question group which in turn asks the candidate a number of questions shown in the bottom right.

<click>

S12.

Our systematic literature review included the following 5 databases of Scopus. Science & Technology, ACM Digital Library, Google Scholar and IEEE Explore

and the used the inclusions of

1. Available online

2. Article is peer reviewed

3. Full text is available in English

4. Article on or after 2005

5. Can be an academic or commercial project

and the exclusions of

1. Non English papers

2. Duplicate studies

3. Magazines, newspapers, websites, podcasts, blogs

The criteria and queries used for each database vary and are included in my CA1 document along with their respective counts through each filtration. Broadly however the results are shown in the following slide.

<click>

S13.

Each filtration reduces the number of suitable research papers. The first filter takes into account any inclusions and exclusions which were just mentioned. Filter 2 then removes studies based on their title review. Filter 3 based on their abstract and finally filter 4 on a broader review of the complete study .

<click>

S14.

the comparative analysis shown here summarises our findings which are explained in further detail within our CA1 document and so I encourage anyone that needs further detail to read that document. The analysis takes our filtered research papers and then asks a number of questions being

Was the research performed on a closed-ended questionnaire?

Were machine learning techniques applied?

What data mining method was primarily used?

Was the dataset a medical dataset?

If association rule mining was applied were dynamic membership functions used?

Were any anomaly detection techniques used to filter out rare cases?

Did the researchers compare questionnaires completed by the same candidate at different times?

Did the research attempt to improve upon the final results through fine tuning using neural networks?

Our contribution is described at the base of this table and attempts to address each of these points

<click>

S15.

From our comparative anaylsis previously shown we have discovered a number of literature limitations or gaps within the research community

<click>

1. Very little research that is able to mine association rules from closed-ended questionnaire data and none that supports medical questionnaires

<click>

2. The work that has been undertaken use static membership functions that are both time consuming and expensive to create

<click>

3. There is a real lack of analysing associations between such questionnaire’s over time from the same individual

<click>

4. No relevant work builds on the accuracy of mined rules through fine tuning of parameters using machine learning techniques

<click>

S16

Working from these gaps we can then propose a number of pertinent questions

1. Is it possible, in a timely manner, to reduce the need for a physical medical assessment for a job role by introducing a suitability predictor using only responses given in a medical questionnaire?

2. Is it possible to improve upon the suitability predictor by allowing actual medical assessment results to be fed back into the live system?

3. Would removing rare or anomalous candidates from the pool of candidates create a better suitability predictor?

4. How to analyse and compare the results of repeat medical assessments from the same candidate for different job roles over time?

And lastly given the previous question our research would add a final question of

5. How to verify and validate the above aims?

<click>

S17

From these questions we can imply the following objectives

1. To classify a candidate into a small number of groups that give a sliding suitability score.

2. To define a mechanism whereby results of physical medical assessments are fed back into the system for a better predictor

3. To build an anomaly detection routine to predict a list of candidates of concern.

4. To build a model whereby assessments maybe compared along a timeline so that assessments taken multiple times maybe analysed.

5. To evaluate the developed artefacts from the previous objectives.

<click>

S18

To describe the methodology we have followed I would like to quote Babbie who stated that research is a ”systematic and orderly approach taken towards the collection and analysis of data so that information can be obtained from those data” A research methodology establishes the framework for research, amongst other things it defines strategy, approach and components for the research.

Over time certain methodologies have been put forward that suite the field of Information Systems. It is Design Study Research or DSR which has been selected as the most suitable of these for this particular work. In DSR ”researchers focus on building some kind of artefact they believe will be useful to a particular stakeholder community. They then evaluate the merits of the artefact in various ways” (Williamson & Johanson (2017, [51])).

Figure 1.3 shows the differences between the classic approach, which creates artefacts to attempt to build or test a theory and the DSR approach which builds artefacts that are useful to certain stakeholders.

S23.

Many research methodologies have been put forward within Design Science Research and each at some point each has attracted there own share of praise and criticism. In fact a number of practitioners advise to simply choose one that fits your research and then adhere to it rigorously.

The specific methodology that this research will undertake is that proposed by Peffers et al. shown in the following 6 steps

<click>

Step 1 – Identify, define, and motivate the focal problem

As previously mentioned in the background of today our industry partner has learnt through experience that candidates may fail at a very late stage of the recruitment process causing potential opportunity losses for both the client and candidate. It is this late failure that brings rise to the core problem of this research. How can a potential candidate be assessed on some medical criteria without involving an actual medical assessment?

<click>

Step 2- Define objectives that a solution to the focal problem must achieve

In order to address the focal problem any potential solution should garner useful information from the candidate’s answers to a preselection questionnaire that they are required to complete. The questions contained within any such questionnaire should take into account the specific role for which the candidate is applying and any typical risks or needs that are associated with that role.

<click>

Step 3. Design and develop the artefact

The developed artefact will initially encompass individual classifications for each role from the associated question groups. The classification will allow for human override for a candidate that has been flagged as borderline. Once sufficient data has been collected the ability to include some transfer or cluster learning would remove the need to train all the roles separately. This however is not the initial goal of the work.

<click>

Step 4. Demonstrate the artefact can be used to help solve the focal problem

The specific example shown earlier shows an example of a job requirement for “lifting heavy weights from floor to waist”.

Of particular note are the edges for heavy weight, hip and flexor. It is on these edges that properties will initially be attached such as the value for the actual weight considered ”heavy” by an expert in the field. As our classifier begins to acquire data it is these properties that will dynamically be altered in order to solve our core problem of classifying a candidate.

<click>

Step 5. Evaluate how well the artefact solves the focal problem

My CA1 document explains in reasonable detail the industry standard way that we will verify the “correctness” of our solution which include measures such as confusion matrix, area under the ROC (receiver operating characteristic) curve and F1 score. Please refer to it for further detail.

<click>

S28

Step 6. Communicate the outcomes of the research

Amongst the outcomes of this research will be the development of a number of novel algorithms to be incorporated into a commercial software product. It is the algorithms that are developed during the design phase that will satisfy the artefact requirement of DSR. The stakeholder community will initially involve the industry partner of the university but will ultimately be useful to anyone dealing with the problem of classifying the answers to closed medical questionnaire data.

Through progressing the research to completion the communication of the outcomes will satisfy the industry partner. The wider research community will become aware of the outcomes through publishing a number of papers at recognised conferences as set out in the research plan at the end of this talk.

<click>

S29

Looking into more detail at our research questions, the following describes the general steps needed for objective 1 of our work.

Step 1 allows us to create distinct linguistic definitions that describe our domain feature in a fuzzy way.

We introduce similarity matrix which allow us to compare two terms fuzzily.

For instance if we had a health rating table as shown here we could describe a Very poor rating to have a similarity of 0.5 when compared to poor.

Step 2 will define our initial membership functions that define the degree of truth of a feature. Here we are considering diastolic blood pressure and deciding which value would be considered high or normal

The 3rd step goes on to select noteworthy association rules using the constructs from step 1 and 2 along with any crisp values.

The basic idea of step 3 is to mine a series of rules based on their calculated support and confidence values.

Unlike traditional association rule mining however our questionnaire’s may contain any of the answer types described by Marshall in the background part of this presentation.

Step 4 will improve upon the discovered rules. The work by Mamuda et al has been shown how to tune the parameters of fuzzy rules using traditional gradient descent. The added advantage of this was that it ”allowed a membership function of the rule to be used more than one time in the fuzzy rule base”.

<click>

S30

The physical medical assessment is outside the core architectural boundaries of the current system. There

are many factors to this including geographic remoteness or the use of third party

independent assessors. In order for any selected candidate to be both accurately

determined and done so in a timely fashion it is vitally important that results of

physical assessments are fed back into the system promptly. In fulfilling this the

research will not only complete a design criteria but will also help to fill another

literature gap. Chen et al. concluded that their design proved to be slow

as the fuzzy membership routines were static. In feeding back this assessment infor-

mation from clinical experts into the system the membership functions themselves

will be altered over time and in doing so become dynamic.

The work of Marasini et al. describes how there are at least 4 main approaches that can be used to define a membership function:

 a ”best” mathematical function

 a probabilistic point of view

 decision-theoretic approach

 axiomatic measurement theory

The example function shown here in step 3 represents a typical interpretation of whether someone is young from the work proposed by Cerioli and Zani.

Briefly the function is saying that anyone less than 15 is definitely considered young and anyone above 25 definitely considered not young and then the range of ages between 15 and 25 represents the transition zone between being young and not young. It is this type of function that we intend to make dynamic over time through the feeding back of data from actual medical assessments.

Cheli and Lemmi propose another type of membership function based on a distribution function and others suggest other frameworks to define membership functions but the exact form of membership function is not of concern at the current stage of our work.

The work of Kostikova et al discusses how we often use membership functions with no reevaluation over time and thus our set inclusion may drift. The work talks about two reasons that may cause this drift , the first being objective reason which comes about by improvements in computer processing speed but it is the second reason of subjective reasoning that is the focus of this work. Subjective reasoning cause drift through changes in internal judgements of experts, peculiarities of the perception of the analyzed objects and the environment in which they are located. We hope it is obvious that concepts such as “heavy”, “light”, “comfortable” and “stress” may be judged differently by different experts and also the same experts over time.

Their work goes into great detail about how these membership functions change over time and I commend anyone interested to refer to their study. In fact it is the aim of this objective to harness the dynamic model from this work. The researchers show that the number of parameters and subject area have no bearing on the final outcome of the result.

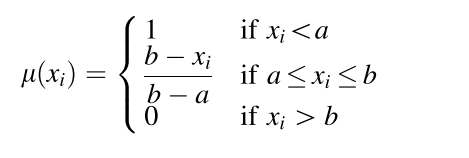
Notes

Fuzzy theory introduces a membership function, expressing the degree of

membership of the elements to a set.

The core of the FS theory (Zadeh 1965)is that the mapping function, indicating the membership of objects to a certain set, can take values on the entire unit interval. In Zadeh’s viewpoint, membership is conceived in terms of degree of inclusion, i.e., partial belonging to a certain set.

It is reasonable to assume that people can be classified along a poverty/non-poverty line through a continuum and gradual membership, and not through a dichotomy. In summary, defining a poverty line leads to uniquely discriminating between poor and non-poor people, that can be considered as an assumption not totally corresponding to reality. A seminal proposal in order to re-consider the poor/non-poor dichotomy has been given in Cerioli and Zani (1990), who suggested to introduce a transition zone between the two states indicating poverty and non- poverty. In their original paper, the authors considered, for each subject i, her/his equalized income xi and defined a transition zone (a, b) into the poverty line. Moreover, it has been suggested that the membership degree can be defined through a linear function:



S 31

Objective 3 involves building an anomaly detection routine to predict a list of candidates of concern,

The steps shown here take us from selecting the anomalous features , through to analysing, cleaning and normalising those features defining the network architecture as well as training options and lastly training the model.

We have to date implemented an anomaly detection technique originally in Octave and later ported into MATLAB which will discussed in our section on progress to date.

The implemented work uses a current production system that has a lot of data but lacks some integrity constraints to ensure that data is always correct. A similar approach will be undertaken with the system

currently being developed once sufficient data becomes available

S32.

Objective 4 which attempts to compare assessments along a timeline will use a lot of the work proposed in recommender systems.

// and in particular the work regarding local reprresentative-based matrix factorization of LRMF

Firstly we Select representative questions from each questionnaire that are able to be

transferred Then we Apply the work of Shi et al to create a model that splits the representative

questions into global and local attributes. Lastly we Deduce whether sufficient local representatives of an unsuccessful candidate from one questionnaire domain match those of a successful candidate from another domain and if so recommend this candidate for the alternate role.

S33

Objective 5 evaluates the artefacts from the previous objectives,

In order to do this we will use a number of publicly available tests to indicate the correctness of our classification and the slide contains a number of those tests. For further detail on this I would suggest referring to our CA1 document.

S34

With regard to the significance of this work there will be both a scientific significance and also a practical significance,

With regard to the scientific significance

We will add to the very sparse representation of machine learning techniques that specifically target closed-data questionnaires.

The research will apply non static membership functions to the developed Association Rule Mining solution through real-time feedback of the assessments from subject matter experts back into the system.

We will develop a hybrid solution whereby any rule parameters that are mined will be further tuned through the use of gradient descent with the aim of thereby demonstrating a better solution.

<click>

S35

The practical significance, from an employers point of view, would include the facts that

1. Medical assessments are a bottle-neck in a candidate’s selection process because of their late placement in any recruitment campaign. This lengthening represents a potential missed opportunity for the candidate of missing other company campaigns.

2. The added physical assessment time is also a potential missed opportunity for an employer as a final decision on a promising candidate may arrive too late as the candidate has already been picked up in an entirely different role. Indeed if this is the only chosen candidate the whole selection process may have to begin again.

3. The cost of physical medical assessments may also prove prohibitive to an employer.

<click>

S36

Our research plan is as follows, as you can see

we have finalised the problem and research questions, we have finalised the literature review, we are in the process of finalising our candidature assessment 1.

For the next year we plan to prepare a conference paper based on research question 1 and prepare for candidature assessment 2. Also we will implement our medical assessor feedback work so as to satisfy the dynamic membership functionality for our research question 2.

Following that year we plan to complete the remaining questions 3-5 that I have mentioned today along with preparing another conference paper for question 4 of our research. After this point we will be completing candidature assessment 3 which will lead into our final goal of the thesis write up.

S37

The progress to date has involved

An initial proof of concept which has been developed solely from an initial system that WHO was already using in order to get an idea of the completeness of the data. At this point in the research we have used both the Linear regression and support vector machine or SVM to detect anomalous candidates. At this point having the ability to exclude rare occurrences was considered more important than the accuracy of those decisions. Moving forward it is the intention of the research to introduce a Deep Neural Network architecture

The current system although lacking integrity does contain sufficient data for a classic split of the data into training, cross validation and test data. When the new system comes on line a less rigid split will need to be used involving some kind of k-fold cross validation to cater for the initial lack of data. To date some work has been done using a leave-one-out technique to gain familiarity with the approach.

We attempted to predict the labels of new data and calculate the classification accuracy. Both linear regression and SVM struggled to correctly classify anomalous candidates. In total from a test population of 2834 there were actually 20 candidates rejected however our solution predicted none of these. Upon closer examination of these 20 however it became apparent that these candidates had been rejected for reasons outside of the questionnaire. For example one candidate was rejected after the assessor discovered that they had only one arm which was a question not asked through the particular job questionnaire. Other candidates were rejected for reasons that the company gave to the assessor directly and not through the questionnaire process. These findings will be fed back into giving the next iteration of the system more data integrity.

As previously mentioned the solution was originally written in Octave and later ported to MATLAB

<click>

S38

In order to verify that fuzzy Association Rule Mining is suitable for our solution the results of a few research papers have been reproduced even though no formal means of reproduction exists.

The research studies have been developed using a neo4j graph database so that the implementation of other studies and indeed our own work should be able to be fast tracked as it is only the data of the work that needs to be created and the subsequent extraction of rules has already been implemented.

<click>

<click quickly and references>

Thanks these are the references I have used and happy to take any questions .