Speaker	Professor Eun-Jung Holden
Talk title	Humans, Machines, Rocks and Data
Venue	Ivy & Jack
Time	Tuesday 29 October 2019, 5:30pm

Rosanna Marchesani

Good Evening Everyone!

I hope you are all enjoying the lovely venue and have had a chance to get yourselves a drink.

Firstly, I would like to acknowledge the Whadjuk Noongar people as the traditional owners of the lands and waters where we are meeting today and pay my respects to the elders, past, present, and emerging.

As you know, tonight, twenty-two academics are speaking in ten bars across Perth. At UWA we are excited to make education of this cities popular culture through city bars into places where you can enjoy a drink, whilst learning about the impact that our research has in the community.

Earlier this year, UWA launched 2030, our vision for the next decade. A significant component of UWA 2030 is to harness and nurture the collective intelligence of our students and researchers to help tackle the grand challenges, facing our society and our environment.

If you are sharing Raising the Bar on social media, please tag, @UWAresearch or #rtbperth19. Tonight's talks are being recorded and will be published as podcasts on our social media channels and those of you have tickets for tonight will receive an email to let you know.

Now, to introduce Eun Jun Holden.

She was trained as a computer scientist at UWA and worked as a researcher specialising in deaf sign language recognition and visualisation using computer vision and graphics techniques. Since her transition to geoscience in 2006, she has developed and commercialised a number of data analytics tools which have had significant uptake by the resources industry globally. Her current research focuses on machine-assisted modelling of geology and resources using interpretable and actionable data science solutions.

Please join me in welcoming EJ to the stage.

[clapping]

Professor Eun Jung Holden

Thank you very much and thank you very much for being here.

I was wondering whether my topic is interesting enough to track down an audience for this talk but it is nice to see your faces here.

As mentioned, I was trained as a Computer Scientist so my real specialty was Image and Automated Image Analysis in which you are learning to identify what a person is actually gesturing in terms of Australian sign language.

Then in 2006 I had an opportunity to change and move to Geophysics and it was an interesting journey, because in 2006 there was a mining boom and the Geophysics Group at UWA just couldn't get a Geophysicist to work in their team and so they decided to get an Image Analysis person instead and that's where I came in.

Now, my journey in Geoscience was that I was very eager to apply automated techniques to diverse types of data but really, what I had trouble with was that there was actually a gap between how the outcomes of automated analysis could be actually used by end-use of Geoscientists because sometimes its generates false positives that actually for them clean up these false positives takes longer than interpreting the data and sometimes the algorithms are so complex that it is very hard to explain or understand how this process works, which actually makes end-user Geoscientists very difficult to make a decision based on these outcomes of automated analysis.

So, going back, in 2006 since I started this journey, one of the workshops that I was in, it was within the first year of my arrival in Geoscience, I was very new, I didn't know anything about Geology or Geophysics much, but this forum was, everyone was a Geoscientist, except myself and the forum was discussing, "What is the future of Geoscience and how do we train our Geoscientists student to be ready for their work in the industry?" and people started talking about the importance of fieldwork and being able to recognise rocks, being able to understand that Geology is associated with this rock identification.

Very naively, they asked my opinion, so I gave it to them and I said, 'One day there will robot Geologists in the future, why do you need to train these Geoscientists to do that in the field?" and then there was burst of laughter in the room and deservedly so because the last thirteen years to appreciate the complexity of Geology.

We are talking about an environment that has had millions and billions of years of geological events such as plate tectonics or earthquakes and the volcanic activities as well as erosions. These are actually such complex events that happen over such a long period of time, so Geology in one location to the other is never the same.

Can you imagine, modelling them is so extremely difficult?

Nevertheless, in recent years, we have seen the surging interest and eagerness by resource industry and wider Geoscience community on how to actually engage and integrate the machine solutions for Geological interpretation of data and this is based on just the increasing volume of data and the diverse types of data that people are collecting and have access to and it is difficult for humans, we have a lack of ability or a limited ability to recognise patterns or retrain information when you are dealing with large volumes of data with a diverse type of information.

The interest by the Geoscience community of looking at the machine solutions to help this task is not a surprise.

You will hear often about, machine learning AI (Artificial Intelligence) and other computation methods that can help Geoscience data analysis.

I just want to give a quick definition. What is Artificial Intelligence? It is a machine that has an intelligence, just like humans. So, being able to visually see your environment and being able to understand the scene and objects within it, like faces, beer and the glasses of wine. Understanding the visual scene and also reading documents and understanding language and being able to hear sound and recognise voices or the speech, these are all human intelligence we have taken for granted and to actually build those systems, machines can do all these tasks, it's not a trivial task and added to the complexity of for Geoscience is that we are actually trying to understand a very complex model which is the Geology and using what's known as SPARS data.

If you look at the minerals industry, you are talking about the most high-density data they have, is probably at a resource evaluation stage where they drill holes which is the most likely ground truth that you have about sub-surface because you drill holes, you take the rocks out and being able to visually see whats on the sub-surface.

When you have this data density environment and resource valuation, the drill holes have a diameter of about 8 cm and it's on a 50m x 50m grid. So, if you calculate the area ration, we are talking about the sample we have is less than 0.002 %.

What we are really now doing is that using SPARS data, we have lots of them, companies drill hundreds of kilometres of drill holes a year. So, you have lots of data but if it's sparsely located spatially and being able to understand this complex geology, what's happening between the holes is actually a task.

Humans do it easily but for a machine, it is very difficult.

Now, going back to my research.

What do I actually do?

My research focuses on how to actually apply these machine solutions that might use machine learning, automated analysis, image analysis as well as interactive visualisations to help people interpret Geology better from diverse types of data sets.

I just mentioned about the drill hole data, we actually are dealing with data sets that are so diverse that even in drill holes, we have geophysical measurements, there might be density, magnet susceptibility or gamma and so on and that is actually measured down the hole. But this geophysics measurement actually can be captured from the surface but also from the air in the airborne surveys and we have geochemical data which are chemical element concentrations when you get a rock sample, you can see these chemical element concentrations to understand the mineralogy within those rocks and we also have observation that geologists do in the field and when they actually get the rock, they actually look at the textures and visually analyse them.

When you map geology there are things that you map, so you have types of rocks you have got to identify and probably the layers of rocks of what you have is a stratigraphy that is important information and then the structures if the crust will break that are used as a conduit for the fluid floor.

That is also a very important feature, geological features that people map.

When you have these diverse types of data that come in from drill holes, surface from the air and some people use satellite imagery as well.

You have these large volumes of data then how do you actually then integrate these machine solutions to help the Geologist to actually interpret them.

What my student did, who completed a PhD in 2015, he had an electronics background from Sri Lanka and when he came to my group, we decided to see how uncertain humans are when they interpret the data.

I think that was a good place for us to start and justify why we need a machine solution.

We had an eye tracker that tracks the eye gaze of the interpreters when they were actually interpreting magnetic geophysics data. Magnetic data, the imagery we use is a map that actually displays the magnetic susceptibility variations of the sub-surface. This particular data set is the porphyry rich area. Porphyry is a rock that is rich in gold and copper and we actually had this porphyry targets that looks like a Mexican hat shape features in this magnetic data.

We thought this was a very easy target for people to recognise because it's a circular feature and it has a quite distinct signature, so we gave this image with lots of porphyries, porphyries actually occur in clusters so you have a cluster of porphyries from top left to bottom right and we gave this data and asked fourteen interpreters to detect where are the porphyry targets and we used the eye tracker to track how they are looking at the data and how they are detecting those targets.

In our experiment, amongst the fourteen interpreters, given the number of true, the targets that we were looking for, the detection rate of those targets by interpreters ranged from about 20% to 76% and the average was about 36%.

Interestingly, when you turn this map 180 degrees around and get them to take the target again, we found significant variations. We did individuals as well.

What it means is that where this visual feature that you are looking for, a target that you are looking for appears, somehow the clustering pattern was impacting how they can actually find these targets. So, we also found out how you enhance these images has an impact on how people are looking and finding these targets and this was actually quite a surprising result for us, considering how easily thought the task was.

However, this target detection rate aligns with another study done by a researcher from England, called Clare Bond. What their group did was, they generated a very simple geological model where you have a simple tectonic setting with three major structures which are the fractures in the sub-surface and the fourth model, the geophysics data.

So we generated a geophysics map that actually is aligned with what is seen in the model and given it to over four hundred interpreters. Only 23% identified the tectonic setting correctly and only 21% detected the structures.

So, these studies really show how variable human interpretation is.

However, we know that we do make good decisions through intuition and through experience at learning and sometimes our presuppositions and our background affects what we see and that is actually in psychology called "inattentional blindness" and there are various studies or the video clips they have seen when people are playing basketball you are asked to the ball passes and then there was gorilla walking past and just don't see it.

You only see what you are looking for. This is actually problematic for geology because really, it is so complex, you don't exactly know what you are looking for.

All this added complexity shows that we need support from machines. For machine support as I briefly mentioned before, that applying machine support or machine analysis on your data is not that useful sometimes if it generates results that doesn't align with your thoughts or your decisions.

For example, we had a project that I worked on at the very beginning after my transition to Geoscience was generating a liniment detection system which is the automated image analysis tool that finds the discontinuity within magnetic data.

What we found was that when you actually, automatically generate these liniments, it is a very nice first-class tool but if you want to use that mapping structures, then you have a lot of cleaning up to do because there are some features that are edges and images but it's not really structures, geologically.

How do you then overcome the discrepancies of automated analysis with geological decisions?

I just want to mention a little about, I just mentioned about Artificial Intelligence before. Machine Learning, probably you hear these words a lot, lately. Machine Learning is an application of AI where a computer program actually learns without explicitly being programmed. What it means is that … let's just use online shopping as an example. You can explicitly provide a rule to provide a suggestion but online shopping when you are doing it, what they doing is learning, all those patterns that people have purchased over time and it is actually learning the associations of combinations of items that they actually buy and using Machine Learning, is learning these patterns and they can provide a suggestion of what you should be buying next.

In my research, I use the Automated Image Analysis and also Machine Learning. The problem with Machine Learning there are simple algorithms that can be explainable but as the algorithm gets more sophisticated, it is very difficult to explain how the system processes the input and then generates the output, the way it is.

When you have a Machine Learning that interpretability is a big issue. If you are a geologist in industry and you ran this Machine Learning system and you got some outcomes and said you want to go to a boss and say, "I want to drill the next hole right there because my Machine Learning algorithms say so." If you don't understand how this Machine Learning works it is very hard to go and sell your decisions.

What really is going is that sophisticated algorithm which performs very, very well, like a Deep Neural Network for example or Hierarchical Neural Networks but because of this difficulty of explaining the actual way of how it works is problematic. But, if you are talking about interpretability of Machine Learning, it's not just understanding how things work, from data, scientist's point of view, someone who implements these things, how the weights change between the connectivity of neurons in the algorithm is important but that has nothing to with you, if you are an end-user. What you interested in as an end-user is that how the variability of your input affects your output and how your control of different parameter settings of your algorithm actually affects your outputs.

These are the questions that the end-user wants to know.

But then, there is another problem with this, actionability. So, if you want to use these Machine Learning outcomes for your decisions, as I mentioned before, you want this solution to be geologically feasible so that when you actually see the result, it doesn't look like, "it can't happen" so to achieve that, it has been an interesting journey for our group, because my research team has been working in this space where we develop and apply algorithms and get the outcomes but we spend equal amounts of time deciding how to make this usable for end-users and I would like to share a journey that we had with our major industry sponsor of our team, Rio Tinto Iron Ore. Actually, we have a representative here as well.

This is the system that automatically classifies or semi-automatically in a machine-assisted[34:53] way to classify stratigraphic sequences in the drill hole data. So, they have a 500 – 600 kilometres of data, drill-hole data to process a year.

Interpreting those at every two-meter intervals and seeing where you are in the stratigraphic sequence, is actually really hard work to do as a human interpreter.

Why Stratigraphic Sequence Analysis is important is that these are, as I said before, stratigraphy is a sequence or layering of rocks because mineralise the layer of this stratigraphic sequence are the ones that are fundamental domains for resource evaluation and also, they are important for how to build a mine pit because you want efficient mining of these holes.

When you actually use this stratigraphic analysis using drill-hole data, what we have done is that we realised that there are three types of data that interpreters use in the environment.

One is human interpreters use what's known as Wireline Gamma which is the Wireline Cabling technology that goes down the hall and it captures your physical measurements but there is a natural gamma which is Gamma Radiation from the rocks that I captured and this data is very important for the iron-ore operation because in the Pilbara there is a different formation which has got these shale bands which is a rock type in the layer and that shale band has distinct signatures associated with this natural gamma data.

If you actually understand that pattern or characteristic and identify where shale is, you actually know where you are in the stratigraphic sequence.

Natural gamma from wireline we had a geochemical data which is … when there is a bag of chips, rock chips comes at every 2-meter intervals, they go to a lab and comes back with multi-eliminates it is a geochemical element concentration values.

We have that … how much time have I got left … thirty minutes and then we also have a geologist log. This is a difficult one because I went to a site where there was this scorching heat of over 40°, there was a geologist standing outside and logging those bags of rock chips coming out of drill holes, every 2-metre intervals and they are logging percentages of mineralogy, rocks and textures.

Now, as you can imagine, the uncertainty that goes with it.

We had those three data sets and tried to actually classify stratigraphic units and the sequence along the depth.

We use some very sophisticated Machine Learning Algorithm called Convolutional Neural Networks which is going to be … we gave up on explaining it, but what we could do is that we sub-modularised our problem so that we built classifiers that optimise for each of the data sets, so when the result output comes out, the geologist can sit down and say, “okay that makes sense”. At least, being able to understand whether this is actually working or not and what is actually a data set which is identifying which stratigraphic units and that validation is actually, to me, a very important one.

We built classifier of three different classifiers for different modality data and we also under that geologists in the company don’t just use those three data sets, they want data that goes to the historic model and see where the boundaries of those stratigraphic units lie and see whether they agree with the historical model and then you also bring in different things like surface geology so they can concentrate on what is the first layer in the stratigraphic sequence and also you bring in geological structures, structure data that might change everything in your interpretation and you also have the thickness, expected thickness of those stratigraphic units and so we need to honour or retain the actual workflows that actually geologists do, otherwise this system is not very usable.

Seamlessly integrating Machine Learning into a geologists workflow and making the whole system usable without making too many changes in what you actually do in a day to day operation, was actually a key to the uptake of this type of system and as I said before, we had the one input data, that was highly uncertain which is the human logging of material types that exist.

The human logging shouldn't be used as is, it has to be somehow validated and we are very lucky because the sponsoring company actually validates every 2-meter interval when the geochemical essays come back from a lab, they actually validate whether the human logging is actually correct.

The way it does it is that because it has a theoretical value that are theoretical essay element concentrations that are associated with each material types that they log and percentages. So if you can actually convert into a theoretical essay and then the lab essay comes in and you can use them to minimise differences by correcting you geologist logs.

So, it is a mathematical optimisation that's what we thought at the beginning but when you run [inaudible 41:58] optimisation, it actually comes up with solutions that are not really acceptable by geologists.

What we decided to do was that collected how top experts validated, actually validate these logs, so how they change different mineralogy types when they validate and what other combination things that exist in their project area.

Going back to online shopping we use the same technique. So what other associations of different mineralogy that can co-exist and how you can actually swap different material types. So, when you do that you do have geologically feasible solutions, you constrain your optimisation using them.

As a result, both of these systems AVA and Meddy are extensively used in the Resource Evaluation Team at Rio Tinto Iron Ore.

For us, as a data science team to see your system actually embedded in operations and being used, there is nothing better than the feeling of reward that you get and seeing the impact of your research.

That was actually an example that I wanted to give about how to make data solutions or Machine Learning in this case, usable or generating geological feasible solutions and while preserving or honouring the workflow that geologists already have and that makes it easier for the uptake of this type of technology.

What else are we doing in our research group?

That was a very much delivery focused project but we also have a research project by PhD students and I think Machine Learning now and in the future has a big place of reducing uncertainty of when you generate the geological interfaces in the sub-surface and the completing surfaces that we only partially know and how to integrate different modality data, capture those different solutions.

How do you actually integrate geophysics into your chemistry to interpolate the spatial distribution of geochemical concentrations for example?

Another project that's worth mentioning is the students who are working on Knowledge Graph. If I remember correctly, when you are talking about AI it's not just seeing and understanding the scene or hearing sound and understanding speech or understanding written text, those are information which is a refined form of data. But, the reasoning that comes in to generate knowledge is another step.

Knowledge is when you actually make meaning out of that information that you see, how you can collectively generate your knowledge. It is actually a very important part of human intelligence and also being able to correct when the new information comes in.

One of my PhD students in computer science collaborating with [inaudible 45:49] experts in computer science. What we are really doing … I will finish this soon … so, what she is doing is using geology context and how to build associations to change geological keywords with a specific aim to understand, what the mineral depositional environment for particular deposits.

For gold mineralisation what is the host rocks of this gold demineralisation? What is the geological era and what is, probably stratigraphic where they appear? Or what is the actual location, spatial location that is special about this place?

This type of knowledge that is generating from Knowledge Graph, she is doing this automatically, generating these associations and what we are hoping is that there is this real problem with mineral exploration at this point of a conceptual model that where this mineralisation outcome seemed to vary from people to people, group to group.

So this type of work will generate a more evidence-based approach to mineral exploration.

Thank you.

[clapping]

Rosanna Marchesani

Everyone, we have a few moments for questions so we would like to open up to the audience.

Audience

[inaudible 47:41]

Professor Eun Jung Holden

I think if we do a take one step at a time because if you are talking about AI geologists, Artificial Intelligent Geologists who can do all the geological modelling using data, I really believe at this point that the project that we have been working so far has demonstrated that Machine Solutions and humans need to work together to get synergy between their strengths.

So, humans who actually have domain knowledge which sometimes can't be explained but it’s there, it comes out when you actually get your intuition, is the thing enhanced by automated analysis results.

Really for the Meddy project that I am talking about, they spent only a fraction of time analysing historical data in their project area. So, what they do now, is that they have time to study geology which is the spatial aspect. What does 3D look like and what does the depositional environment look like for this kind of deposit.

Audience

Thanks for a great Intra disciplinarian presentation and you are a great Intra disciplinarian.

Now, you raised and issue with Machine Learning, that the machine figures out things that people never figured out before.

Professor Eun Jung Holden

Sorry?

Audience

The machine will figure out ways of doing things and solutions that people hadn’t figured out before and now I recall whether it is ENMAS or other areas, if I figured out a new way of doing something, the teachers would say, “Tell me how you did that?” “What does it mean?”

You mentioned that from this Machine Learning, we can't quite figure out how they did it, so that must be a problem across the whole space.

So whats happening in that space?

Professor Eun Jung Holden

That is a big problem of explainability of AI. So, if you got to a top Computer Science Conference, that’s still a very much active area of research because computer scientists realise that really, the decision [inaudible 49:54] is very hard to develop if you can't explain how the algorism process works but on the other hand, in our experience like the example gave you, if you sub-modulise your problem enough, that when you actually get the results, understand what is working and what is not, that gives you the confidence to use it.

Rosanna Marchesani

Do we have any more questions from the audience?

Then could we all please … oh sorry!

Audience

From your experiment and you ask the geologist in the pretext and he will say 30% or it's something, he will say [inaudible 50:51] I mean, you say … it takes a long time to clear the false-positive, so I mean ….

Professor Eun Jung Holden

Okay, so I think it's kind of mixed up here but I think the first one that I was talking about is the porphyry detection of magnetic maps. That has nothing to do with our project as I was talking about before.

That project actually, we developed an automated system to identify these anomalies in magnetic maps which actually later that was commercialised by Geosoft Oasis Montaj plugin and Geosoft is now marketing that product and that has been selling for the last ten years.

We ran that and of course, there are false-positives but really, the performance by humans is missing out on the true positives and automated systems generating false-positives is actually, I don't know, I am not a geoscientist so how should they actually use that in industry.

We were given that information by this company for use of porphyry global exploration about ten years ago …no, eight years ago, that they found this channel cluster in Chile and this tool that we developed was contributed to them finding it and only four months ago … two months ago I was in Vancouver and another group actually taught me that they actually found a deposit which was confirmed by this tool.

I am just thinking … I mean financially that plugin wasn't selling very well because we actually commercialised three systems and that's the least performing one, financially, because there are not that many companies who is doing porphyry detection.

But, porphyry exploration, but what it means to me is that the people who use these tools actually can use it as a first-class analysis.

You don't have to take everything as truth and map it, but if you run this type of filter and identify some spots and then start to go in and explore.

I think the company that I was talking to was using it that way.

Rosanna Marchesani

Do we have any more questions?

No? If not we would like to say thank you to EJ for a great talk.

[clapping]

We hope you have all learned something new.

If you are staying for the next talk, we would ask that you please let the girls at the door know that you need to just show again for the next talk because we do have another full house.

If you would like to stay but you haven't registered, then also please let the ladies know at the door and they will see what they can do to let you in.

Thanks so much, everyone have a great evening.