Investigating User Confidence for Uncertainty Presentation in Predictive Decision Making

Syed Z. Arshad[1,2], Jianlong Zhou[2], Constant Bridon[2], Fang Chen[2], Yang Wang[2]

[1]School of Computer Science & Engineering
University of New South Wales
Kensington, Sydney
sarshad@cse.unsw.edu.au

[2]NICTA
Australian Technology Park
Eveleigh, NSW
{syed.arshad, jianlong.zhou, constant.bridon, fang.chen, yang.wang} @nicta.com.au

ABSTRACT

Machine Learning (ML) based decision support systems are often like a black box to non-expert users. Here user's *confidence* becomes critical for effective decision making and maintaining trust in the system. We find that user confidence varies significantly depending on supplementary material presented on screen. We investigate change in user confidence (in the context of ML based decision making) by varying level of uncertainty presented (in an online water-pipe failure prediction case study) and find that all 26 subjects rated higher uncertainty task to be most difficult and had lowest user confidence in predictive decisions of the same. This agrees with our expectation that increased uncertainty would reduce user confidence in predictive decision making. However, ML-researchers subgroup reported being most confident when uncertainty with known probability was presented, whereas other subgroups (viz. general staff and non-ML researchers) appeared most confident when uncertainty was not at all presented. This is an original research to improve understanding of user's decision making confidence with respect to uncertainty presented in machine learning context.

Author Keywords

User Confidence, Uncertainty Presentation, User Uncertainty, Model Uncertainty, Decision Making.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Living in the age of big data and predictive analytics, we continuously find ourselves coming across machine learning based appealing viewgraphs and other (handy) predictions that seem to work (or have worked) surprisingly well in practical scenarios (e.g. the remarkably successful US Election predictions by NYT FiveThirtyEight Blog). So far these machine learning success stories originate from technical experts like Nate Silver (statistical learning) or computing professionals (e.g. self-driving Google cars). But this popularity of machine learning and predictive analytics has created a growing demand for similar tools in non-computing communities as well. People with no background in machine learning or data analytics, would also like to use these powerful techniques and algorithmic variations to their benefit. However, for many of these non-technical users, machine learning based predictive analytics software is like a black box, to which they simply provide their source data and (after selecting some menu options on screen) colourful viewgraphs and/or recommendations are displayed as output. How trustworthy is this output? Or how uncertainties were handled/manipulated) by underlying algorithmic procedures is neither clear nor well understood.

A user might be risking too much by completely ignoring alternate scenarios and having complete faith in system output/recommended values. On the other hand, trivializing the recommended values or having low confidence in systems analytics could possibly be wastage of the incredible potential of predictive analytics and machine learning. We think that by (a) making transparent the uncertainties inherent in ML procedures and (b) optimizing the way probabilistic results are presented - we can improve user's decision making confidence and greatly benefit the usage of ML style predictive analytics applications by non-expert users.

Part of our motivation is supported by Winkler (2015) who greatly emphasizes the importance of communicating uncertainties in forecasts (as imprecision and uncertainty are unavoidable in predictive analytics). He believes that making rational decisions greatly necessitates the consideration of uncertainty; and that probability should ideally be communicated in probabilistic form, as it is the mathematical language of uncertainty. He argues, with the help of several effective examples (from different fields), that probabilities are needed to understand the risk associated with potential decisions as well as to determine measures such as expected payoffs and expected utilities. Probabilities are an essential input to decision modelling and decision

OzCHI '15 , December 07 - 10 2015, Melbourne, VIC, Australia

http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx.

making. LeCleric and Josyln (2015) successfully demonstrate that adding a probabilistic uncertainty estimate in public weather forecasts improved both decision quality and compliance (to evacuation instructions in cases of severe weather threats).

Joslyn et al (2013) agree that people need explicit uncertainty information to make better individualized decisions, but emphasize that care must be taken in how uncertainty is presented as it has long been known (e.g. Kahneman et al 1979; Tversky et al 1974) that people make a wide variety of errors when making decisions (or solving problems) involving probability. This could be due to less than professional (or missing altogether) understanding of the probability theory involved.

At a practical level, 'uncertainty' can refer to an *interval* within which the true value of a measured quantity would lie with a given probability and this leads to predictions with limited precision. Since machine learning typically takes place on historical data sets, the models learnt are generally displayed using narrow success/failure rate lines that rarely do justice to probabilistic manipulation and uncertainty inherent within displayed results. Here we consider a water pipe failure prediction case study and by depicting uncertainty intervals graphically - we hope to give the decision making user a better sense of the imprecision inherent in system forecasted results. To understand better how this affects predictive decision making, we differentiate between user uncertainty and user confidence. This distinction was maintained by Patterson and Pitz (1988) who studied impact of information on user confidence and uncertainty. When making predictions about an unknown quantity, a person's beliefs about possible values for the quantity are termed as user's 'uncertainty', while the belief that a given prediction is correct is referred to as user's 'confidence'. They found that user confidence and user uncertainty were affected in different ways by increasing availability of information. User uncertainty increased when the number of different predictions that could be generated with increasing information also increased. However, user confidence decreased when the task became more difficult with increasing information. It is with this understanding of user confidence that we approach user uncertainty and decision making.

User Uncertainty and Decision Making

Making decisions is one of the most complex cognitive processes and has a long history of investigation in different domain areas. For example, Morgado et al (2015) reviewed the impact of stress in decision making in the context of uncertainty and found that this complex cognitive process involves several sequential steps including analysis of internal and external states, valuation of different options available and action selection. Making good decisions implies an estimate not only of the value and the likelihood of each option but also of the costs and efforts implied in obtaining it. In HCI, attempts have been made to enhance intelligent user interfaces by adding measurable decision making capacity to them (Zhou et al 2015). This was done using galvanic skin response (GSR) and pupillary analysis.

Here we focus mainly on the role of uncertainty and how it impacts decision making in machine learning predictive analytics.

'Uncertainty' continues to be defined in many ways for many audiences. For a user, it can be a psychological state in which the decision maker lacks knowledge about what outcome will follow from which choice. This aspect of uncertainty considered by both economists and neuroscientists is popularly known as 'risk'. Risk refers to situations with a known distribution of possible outcomes. This is the type of uncertainty with known probabilities (Platt and Huettel 2008). However, 'ambiguity' is the other kind of uncertainty, where outcomes have unknown probabilities and research in neurosciences (Huettel et al 2006) indicates that decision making under ambiguity does not represent a special, more complex case of risky decision making; instead, the two forms of uncertainty are supported by distinct neural mechanisms.

As per subjective expected utility (SEU), the decision weights people attach to events are their beliefs about the likelihood of events. Earlier on, as summarized by Camerer and Weber (1992), it was thought that people prefer to bet on events they know more about, even when their beliefs are held constant i.e. they are averse to ambiguity, or uncertainty about probability. However, this was shown to be otherwise by Hsu et al (2005) in their study of neural systems responding to degrees of uncertainty. Their experiments and corresponding neurological observations showed that many people are more willing to bet on risky outcomes than ambiguous ones, holding the judged probability of outcomes constant. These findings motivate us to account for both risk (i.e. uncertainty due to known probabilities) and ambiguity (i.e. uncertainty due to unknown probabilities) while investigating variations in user confidence due to uncertainty.

Presenting Model Uncertainty

In this section we look at uncertainty due to data and issues related to its graphical presentation.

As argued by Winkler (2015) probability remains the language of uncertainty. Joslyn and Nichols (2009) investigated if uncertainty expressed as a frequency (e.g. 1 out of 5 or 20 out of 100) would be easier for non-experts to understand as compared to uncertainty expressed as probability (e.g. 20% or 0.2). They used three different expressions for the test and found that all participants better understood the forecast when it was presented in a probability format rather than a frequency one.

Also several investigations have (historically) been carried out to understand better ways of presenting uncertainty inherent in data. One earlier effort includes that of Ibrekk and Morgan (1987) who investigated graphical communication of uncertain quantities to well-educated semi and non-technical people. They used nine displays (including probability and cumulative density) for communicating uncertain estimates for the value of a single variable in an experiment and found that

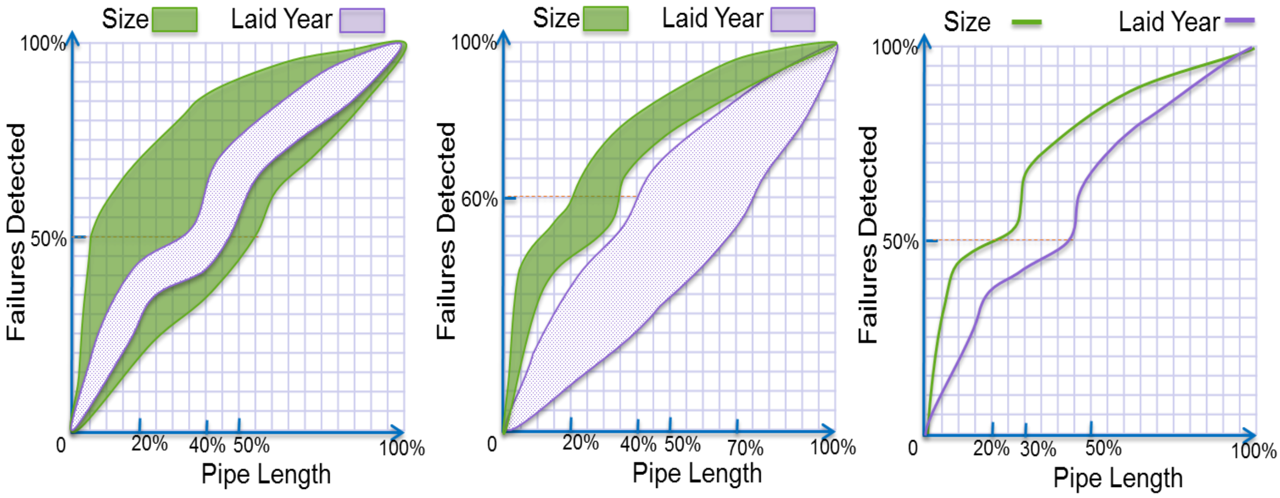

Figure 1. (a) Overlapping Uncertainty (b) Non-overlapping Uncertainty (c) No Uncertainty presented

cumulative distribution functions alone can be severely misleading in mean estimation. Back then it was suggested that future studies of the problem of communicating uncertainty should not focus just on the problems of communicating to 'semi technical and lay people' (as experienced technical people were also prone to making error under certain conditions). We pay heed to this advice and the subject groups we consider in our experiment involve both experts (machine learning & non-machine learning) and general staff.

Furthermore, cognitive load is also known to affect people's use of graphical displays. Allen et al (2014) studied the potential of five graphical displays to communicate uncertainty information when end users were under cognitive load. Results indicated that load manipulation did not have an overall effect on derivation of information from the graphs but did suppress the ability to optimize behavioural choices based on the graph. Overall, the research suggested that interpreting basic characteristics (like 'point reading') of uncertainty data is unharmed under conditions of limited cognitive resources whereas more deliberative processing, like synthesizing information, is negatively affected. 'Synthesizing' here corresponds to cognitive capacity demanding Type 2 processing as explained by Stanovich et al (2012). Type 2 processing is further discussed later in this paper. In our current design we maintain cognitive load to be same over all experimental conditions and keep sessions times minimal. We do this because Arshad et al (2015) have shown that individual cognitive load can vary over longer periods of activity which is then reflected in changing behavioural patterns.

Modelling Uncertainty for Decision Values

In machine learning, uncertainty can be traced to many sources ranging from input values to nature of model representation to final output decision values. Here we concern ourselves mostly with uncertainty associated with output decision values.

Machine learning typically uses historical data sets to learn features/parameters for building models. Users may choose a model that satisfies their personal preferences. Sometimes single feature models can prove to be reasonably good predictors compared to more complicated multi-feature models which could be processing intensive and difficult to maintain. There is uncertainty associated with every feature that is used to formulate the model. Single feature models can be conveniently displayed using narrow success/failure rate lines (See Figure 1c). However, these narrow lines rarely do justice to the probabilistic manipulation and uncertainty inherent within displayed results. Enhancing the thickness of these lines and making it correspond to the uncertainty at given point (i.e. 95% confidence interval) results in scenarios of overlapping (see Figure 1a) and non-overlapping uncertainty (see Figure 1b).

This approach of presenting uncertainty, inherent in machine learning, is merely a beginning to the actual complex nature of uncertainty involved. Here we begin investigations with uncertainty of single feature models and then hope to explore more complex models in future.

EXPERIMENT

Water pipe failure prediction is used as case study for this research. Water supply networks constitute one of the most crucial and valuable urban assets. The combination of growing population and aging pipe network requires water utility companies to develop an advanced risk management strategy for maintaining their distribution systems in a financially viable manner (Li et al 2014). If high-risk pipes can be identified before failure onset, it is likely that repairs can be completed with minimal service interruption, water loss and negative reputational community impacts. Identifying an accurate predictive measure for 'imminent failure' allows utility companies to prioritize preventive repairs that would be significantly lower than the cost of full-scale failures. Thus, utility companies use outcomes from failure prediction models, to make renewal plans based on risk levels of pipes and also reasonable budget plans for the pipe maintenance. Here 'uncertainty' simply refers to an interval within which the true value of a measured quantity would lie with a given probability.

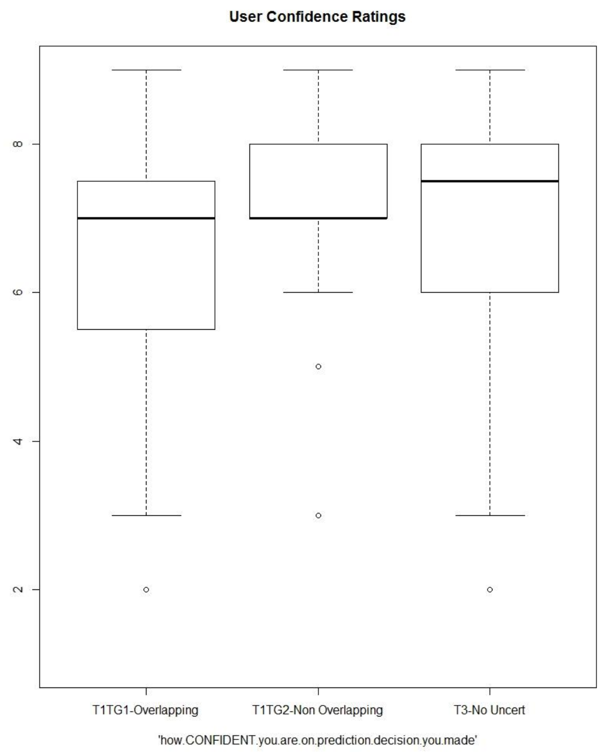

Figure 2. (a) Confidence Ratings

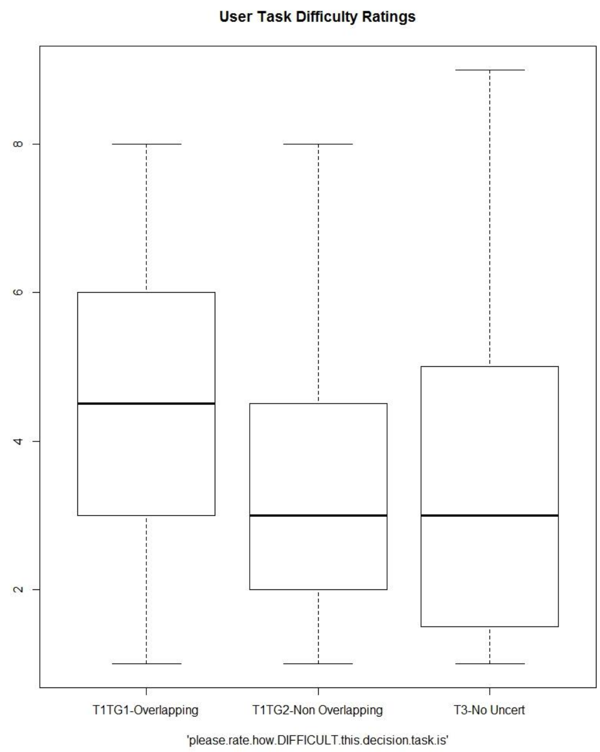

Figure 2. (b) Task Difficulty Ratings

Water pipe failure prediction uses pipe failure historical data to predict future failure rate (Li et al 2014). The data contains failure records of water pipes in a given region. It also includes various attributes of water pipes, such as identification number, laid year, length, material, diameter size, location, protective coating, surrounding soil type, etc. For this research, several failure prediction models were set up based on pipe features (e.g. size, material, laid year etc.). The uncertainty of model output performance was displayed as shaded area (see Figure 1a & 1b) over the actual thin prediction line (see Figure 1c) for each model. The correlation between pipe features and the failure rate from historical data was also recorded. Actual historical data was repeatedly sampled and customized for the simulation of this experiment.

The tasks were divided into three groups (Task Group 1, Task Group 2 and Control Task) based on configurations of task conditions. All models in the given Task group and Control had exactly the same viewgraph with only feature labels changed. The order of tasks and task groups were randomized. The nature of task was on-screen budget estimation with expected variation to be noted as upper and lower limits.

All together there were 26 subjects (each one a user of ML predictive systems at local water department). Ages ranged from 23 to 45 with an average age of about 30 years. Educational qualifications were largely postgraduate (13 PhD, 6 Masters, 4 Bachelors, 3 other). Subject subgroups comprised of nine (9) machine learning experts, eight (8) non-machine learning experts and nine (9) administrative staff.

RESULTS

Box-and-whisker plots for user confidence and user task difficulty ratings are shown in Figure 2. Three conditions, namely, overlapping uncertainty (T1TG1), non-overlapping uncertainty (T1TG2) and no uncertainty (T3) are presented.

A clear trend of lower user confidence can be seen for overlapping uncertainty (Fig. 2a), matched by corresponding values of high task difficulty (Fig. 2b). Using Kruskal-Wallis test for three samples we get p-values = 0.00279 (for user confidence rating) and p-value = 0.0009615 (for task difficulty). With 95% confidence we can say that the respective population distributions are not identical without assuming them to follow the normal distribution.

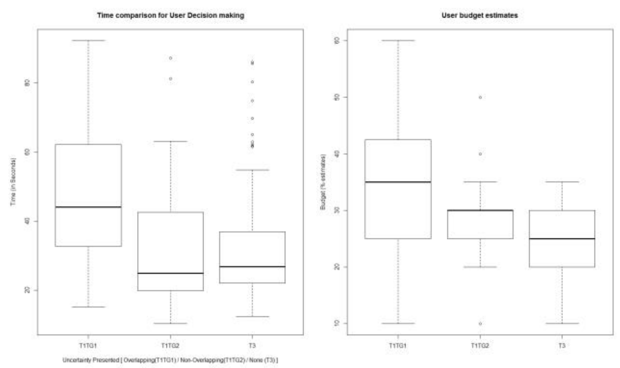

Figure 3. Decision Time & Estimated value comparison

Box-and-whisker plots for decision time and budget estimate are shown in Figure 3. Three conditions of overlapping uncertainty (T1TG1), non-overlapping

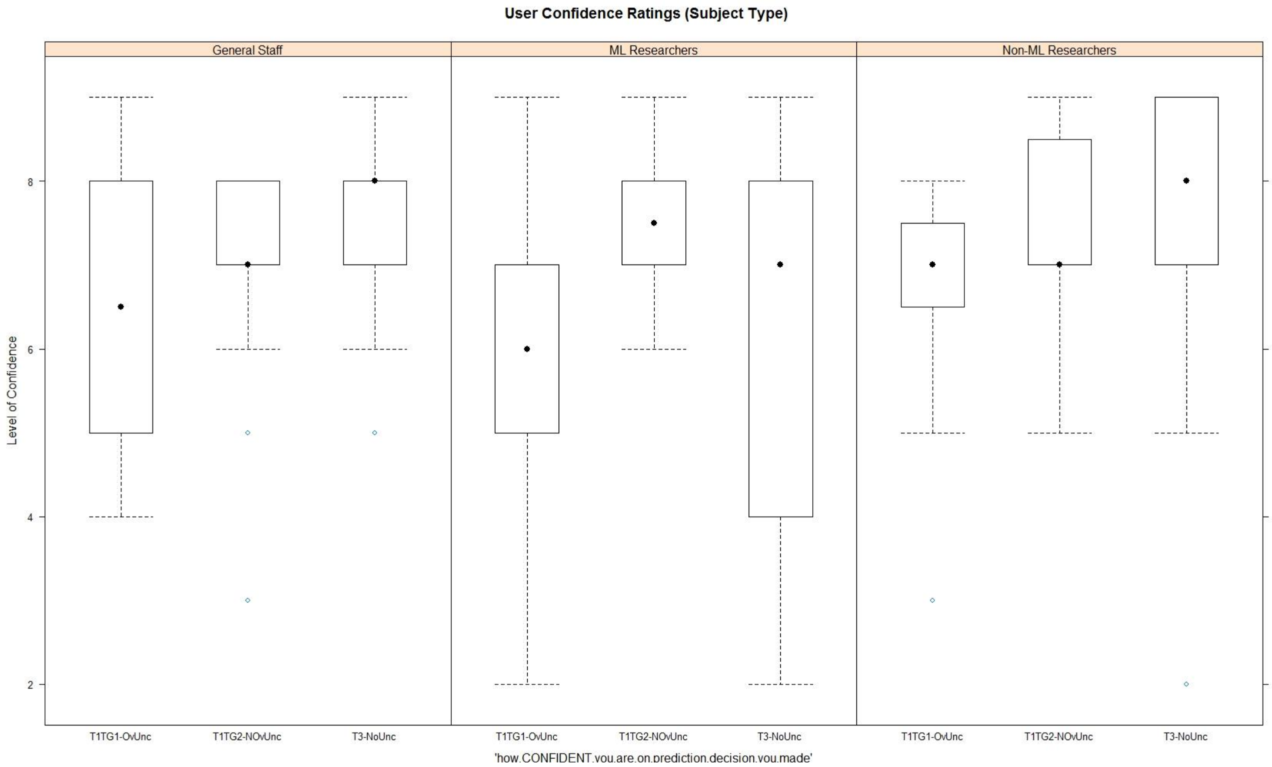

Figure 4. User Confidence Rating (Subject Groups)

uncertainty (T1TG2) and no uncertainty (T3) are presented. Increased decision making times can be observed for overlapping uncertainty task. Also that the budget estimates for the overlapping uncertainty task appear to vary widely, whereas the budget estimates for the non-overlapping task appear quite focussed.

Using Kruskal-Wallis test for three samples we get p-value=1.154e-06 (for decision time) and p-value=3.039e-08 (for budget estimates). With 95% confidence we can say that respective population distributions are not identical without assuming them to follow the normal distribution.

DISCUSSION

Here in discussion we first look at the possible interpretation of overlapping and non-overlapping uncertainty presentations and then analyze the behavior and responses of each subject group accordingly. This is followed by a discussion regarding some experimental limitations regarding user confidence theoretical framework. Finally, we look at some implications of this research for decision making and trust between human and machine systems.

Overlapping & Non-overlapping Uncertainty Presentations

We are already familiar with a useful distinction that uncertainty with known probabilities is popularly known as risk while uncertainty with unknown probability is referred to as ambiguity. By maintaining this distinction we can benefit from all the research that has previously been done with regards to uncertainty as risk and uncertainty as ambiguity.

Referring to Figure 1b, we see that uncertainty can be clearly marked for each model (at all points) as it is non-overlapping. This results in scenarios of clearly known probabilities that make model selection very easy. Simply pick a model that detects more faults for least pipe length checked. This is an easy decision as there is no overlap in models. After model selection, budget estimate is simply a matter of point reading wherever steepest curve tops (the optimum point). Uncertainty in user budget estimate can be matched to model uncertainty depicted by thickness of the line at that point. This type of uncertainty presentation is very close to risk. From a cognitive perspective, making use of this visual presentation should involve more of Type 1 processes that are less demanding of cognitive capacity, holistic, automatic and relatively fast (Stanovich et al 2012).

On the other hand, overlapping uncertainty presentation depicts thick overlapping model lines (Figure 1a). It is no longer clear which model would be the better choice. Much of the decision now depends on the preferences of the user. In case of complete overlap, the model with greater potential gains also is associated with greater potential losses, whereas the model in the middle is more modest in both regards. Trying to make use of this visual, the decision maker comes across ambiguous patches where probability is not exactly known for given particular points. Budget estimation is no longer a point reading task. Thus the user must make some mental valuations of the ambiguous visual and then decide accordingly. We consider this overlapping uncertainty very close to ambiguous uncertainty. Again from a cognitive perspective, making use of this visual presentation would involve more of Type 2 processes that are more demanding of cognitive capacity, analytic, controlled and relatively slow (Stanovich et al 2012).

Accepting Uncertainty and its Impact

In response to a separate questionnaire, all subjects had unanimously agreed to the usefulness of presenting uncertainty as supplementary material. They also agreed to the helpful role of uncertainty whenever it was presented. However, when uncertainty was actually presented, the user confidence generally decreased and the decision making time significantly increased for cases of overlapping uncertainty. Only for non-overlapping uncertainty presentation did time reduce for some subjects as compared to the case of no uncertainty presentation.

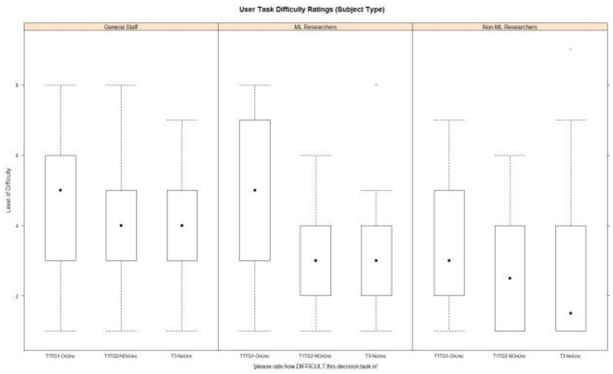

Figure 5. Task Difficulty Ratings (Subject Groups)

To understand this peculiar trend we split up the data to investigate tendencies in each subject group. Group wise data is presented in figures 4 to 7. From left to right, the panels depict general administrative staff (panel 1), machine learning experts/researchers (panel 2) and non-machine learning experts/researchers (panel 3).

General Staff

Clearly the general staff seemed to be most confident when no uncertainty was presented, then became slightly less confident with non-overlapping uncertainty and finally least confident with overlapping uncertainty scenario (Fig 4, panel 1). This is validated by task difficulty ratings that show general staff found the overlapping uncertainty task to be most difficult (Fig 5, panel 1). It also shows up in user decision making time where general staff seems to take the longest time for overlapping uncertainty tasks (Fig 6, panel 1). Finally this results in large varied budget estimations (Fig 7, panel 1), which can be evidence of general staff not being sure or confident. A possible explanation for this trend could be that although general staff may (in principle) think uncertainty can be useful in predictive decision making, yet they are not well versed in its usage when uncertainty is actually presented. This would conform to previous researches of Kahneman et al (1979) and Tversky et al (1974) that suggested people make a variety of errors when making decisions (or solving problems) involving probability. On the other hand this could also be due to these visuals not being the most transparent or user friendly ways of presenting uncertainty to general staff.

Machine Learning (ML) experts/researchers

Machine Learning experts are the only subject group that seem most confident with non-overlapping uncertainty (Fig 4, panel 2). This is actually the right attitude as non-overlapping uncertainty scenario clearly communicates the inherent uncertainty - making the choice of model very easy and budget estimation just a matter of point reading. This is also matched by the lowest decision making time for non-overlapping uncertainty scenarios (Fig 6, panel 2). Clearly the machine learning experts had the best professional understanding of probability and recognized the relevant non-overlapping uncertainty tasks to be the easiest, as evidenced by completion in least time (Fig 6, panel 2) and with most confidence. These observations are as expected. From a cognitive

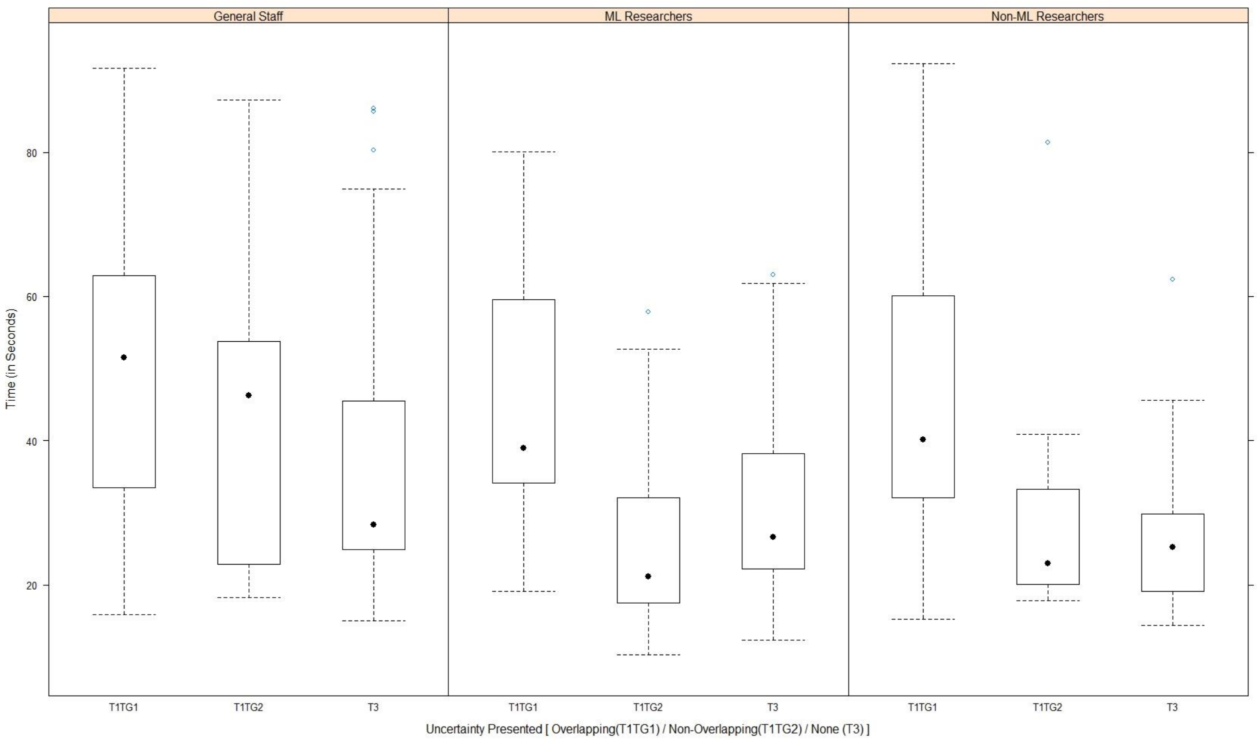

Figure 6. User Decision Making Time (Subject Groups)

perspective, for ML experts this is a case of known probabilities and they feel confident taking up the risk.

Non-Machine Learning experts/researchers
Non-machine learning experts seem to fall somewhere between the two subject groups of general staff and machine learning experts. Most non-ML experts find overlapping uncertainty task to be more difficult (Fig 5, panel 3) and report least confidence (Fig 4, panel 3) when making corresponding decisions in it. Also that overlapping uncertainty task is the case which takes them the longest time to complete (Fig 6, panel 3). These observations can be interpreted to be in line with overlapping uncertainty presentation being viewed as ambiguous uncertainty. However, not all of these non-machine learning experts seem very profound in their perception of probabilistic uncertainty via these visual types. Some of them reported most confidence when no uncertainty was presented (Fig 4, panel 3) and coupled with focused budget estimates (Fig 7, panel 3) – this can possibly be interpreted as a tendency to over rely on learning models.

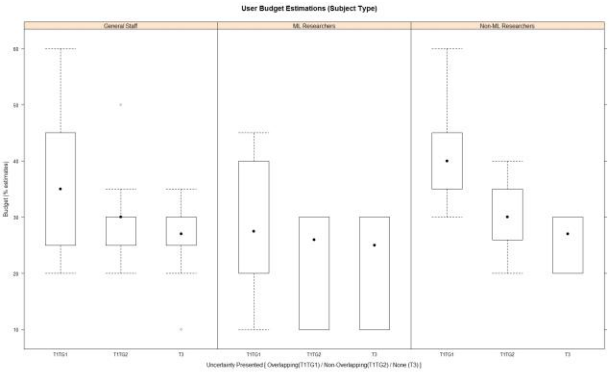

Figure 7. Use Budget estimation (Subject Groups)

Overall, we can say that only the machine learning experts probably understood and benefitted from the presentation of uncertainty in these overlapping and non-overlapping type visuals. General staff and Non-machine learning experts may have agreed to uncertainty presentation being useful, but were not able to fully benefit from what was being communicated. More investigation needs to be done as to whether the general understanding of probability is a factor or perhaps a better form of visual presentation is needed.

Some Limitations and Implications

In this section we discuss some limitations regarding user confidence framework and then implications of this research.

As mentioned earlier, we rely conceptually on ‘user confidence’ ideas, as posited by Patterson and Pitz (1988), mainly because the study addressed user confidence and user uncertainty in the context of increasing information that was critically relevant to our investigations. However, there are several developments (in theoretical constructs of ‘user confidence’) like role of individual differences in accuracy of confidence judgements (Pallier et al 2002), self-consistency model of subjective confidence (Koriat 2012) and collapsing confidence boundary model (Moran et al 2015) that must be acknowledged. Here we justify briefly the relevance of our experimental design in the face of some of these developments.

Conceptual issues concerning User Confidence
Generally speaking, quantitative studies of decision making have traditionally been based on three key behavioral measures, namely, accuracy, response time and confidence. Here, accuracy deals with effectiveness of the decision task at hand. Response time (RT) is the time needed to reach/commit to a decision after the problem stimulus has been presented. Finally, confidence is the user’s degree of belief, prior to feedback, that the decision reached is correct.

In a typical decision making scenario, once the problem scenario along with supplementary material is presented, several other factors can come into play as well. One such group of factors is individual differences that were investigated by Pallier et al (2002). Differences in experience, motivation, attitudinal predispositions etc. can have an impact on decision making process. However, such differences were minimized, in our case, as we categorized the subjects into professionally skilled groups relevant to predictive task at hand.

Next, is the evidence gathering phase, which continues until the user feels comfortable enough to commit to a decision. It is here that studies have shown that the choice certainty (or confidence) is not just influenced by evidence presented (or being gathered) but also by (decision) response time. Greater the time it takes to reach a decision, lower the confidence in that decision (Kiani 2014). We avoided unlimited RT complications by having several similar decision making scenarios administered repeatedly with a soft encouragement for quick predictive decisions once the basic predictive scenario was understood (in practice sessions). Only items that changed in actual testing procedure were supplementary viewgraph types and corresponding data values.

Moran et al (2015) argue that a critical property of decision-confidence is its positive resolution i.e. the positive correlation between confidence and decision correctness. In other words, with higher confidence the decider is more likely to be correct in his or her decision. This is an important component of the dual-stage theories about decision making. Here the key idea is that confidence is affected by the novel information that is collected after the decision is made and its correctness evaluated. This establishes a role for RT2 (the duration of post-choice integration stage) as a prime dependent variable that dual-stage theories of confidence take into account. Moran et al (2015) present a list of empirical patterns involving RT2 that guides theorizing about confidence judgements using their collapsing confidence boundary (CCB) model. Our experimental design does not involve RT2 as the correctness of predictive decisions is not immediately displayed. Also that since decision confidence ratings are collected immediately (once decision is committed by user), our case is closer to that of single stage theories where decision confidence is based on the same information that underlies the actual decision.

Thus we feel that we have sufficiently tailored the process so that the decision-confidence now derives largely from the process of evidence gathering using various viewgraph types. In experiments elsewhere with evidence gathering, the user confidence grows to a certain level before decisions are reached. However, in our case predictive decisions are made in reasonably limited time and thus allowing decision confidence to vary depending on how clearly the evidence is presented/understood in that session.

Implication for Uncertainty presentation efforts in ML

As mentioned earlier, uncertainty in ML arises from many sources. Here we have focused on uncertainty presentation of final decision values only. Several other visual types can be developed to depict uncertainty arising out of input values and also model parameters. However, as we have come to realize from this experiment, the challenge is to communicate uncertainty with known probabilities (risk uncertainty) as that can be the key to higher user confidence (at least for those users familiar with ML). Whereas presentation of uncertainty with unknown probabilities (ambiguity) will most likely reduce user confidence further for all types of users. Also that, users of ML uncertainty visuals can benefit greatly from short tutorials that explain basics of the probabilistic information being communicated.

Implication for Trust research

One implication of user confidence research is the trust between user and the application software running on machine. Since decision making is increasingly becoming computer aided in all walks of life, professionals and experts place great trust in ML-type automated decision support applications that help them get through their tasks. Clearly, an important component of this trust is user confidence; and we now believe that user confidence has a delicate working relation with user uncertainty and presented model uncertainty during decision making. Generally speaking, user friendly uncertainty visuals would lead to quicker and better user understanding. This could be reflected by improved decision making with high user confidence, which can then be expected to strengthen the trust relation between the user and decision aid systems.

Towards the end, we would like to quote Platt & Huettel (2008) who wrote that ‘only through explicit interdisciplinary, multi-methodological and theoretically integrative research will the current plethora of perspectives coalesce into a single descriptive, predictive theory of risk-sensitive decision making under uncertainty’. We feel quite the same about development of effective uncertainty visuals that would help in ML-type decision support applications.

CONCLUSIONS

We conclude that presenting uncertainty on-screen as supplementary material for decision making can significantly improve user confidence. However, care must be taken as to how uncertainty is actually presented. Uncertainty with unknown probabilities (i.e. ambiguity) decreases user confidence, whereas uncertainty with known probabilities can increase user confidence but only if users are well trained in understanding the probability being communicated in those particular visuals.

ACKNOWLEDGMENTS

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

This work was supported in part by the Asian Office of Aerospace Research & Development (AOARD) under grant No. FA2386-14-1-0022 AOARD 134131.

REFERENCES

Allen, P. M., Edwards, J. A., Snyder, F. J., Makinson, K. A., & Hamby, D. M. (2014) The Effect of Cognitive Load on Decision Making with Graphically Displayed Uncertainty Information. Risk Analysis, 34(8), 1495-1505.

Arshad, S., Wang, Y., & Chen, F. (2015). Interactive Mouse Stream as Real-Time Indicator of User's Cognitive Load. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1025-1030). ACM.

Camerer, C., & Weber, M., (1992) Recent developments in modeling preferences: Uncertainty and ambiguity. Journal of risk and uncertainty, 5(4), 325-370.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005) Neural systems responding to degrees of uncertainty in human decision-making. Science, 310(5754), 1680-1683.

Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T., & Platt, M. L. (2006) Neural signatures of economic preferences for risk and ambiguity. Neuron, 49(5), 765-775.

Ibrekk, H., & Morgan, M. G. (1992) Graphical communication of uncertain quantities to nontechnical people. Risk analysis, 7(4), 519-529.

Joslyn, S., & LeClerc, J. (2013) Decisions with Uncertainty: the glass half full. Current Directions in Psychological Science, 22(4), 308-315.

Joslyn, S. L., & Nichols, R. M. (2009). Probability or frequency? Expressing forecast uncertainty in public weather forecasts. *Meteorological Applications*,*16*(3), 309-314.

Kahneman, D., & Tversky, A. (1979) Prospect theory: An analysis of decision under risk. Econometrica, 47, 263–292.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329-1342.

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological review*, *119*(1), 80.

LeClerc, J., and Joslyn, S. (2015) The Cry Wolf Effect and Weather-Related Decision Making. Risk analysis 35, no. 3: 385-395.

Li, Z., Zhang, B., Wang, Y., Chen, F., Taib, R., Whiffin, V., & Wang, Y. (2014) Water Pipe Condition Assessment: A Hierarchical Beta Process Approach for Sparse Incident Data. Machine Learning, 95(1), 11–26.

Morgado, P., Sousa, N., & Cerqueira, J. J. (2014) The impact of stress in decision making in the context of uncertainty. Journal of Neuroscience Research.

Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive psychology*, *78*, 99-147.

Peterson, D. K., & Pitz, G. F. (1988) Confidence, uncertainty, and the use of information. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(1), 85.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *The Journal of General Psychology*, *129*(3), 257-299.

Platt, M. L., & Huettel, S. A. (2008) Risky business: the neuroeconomics of decision making under uncertainty. Nature neuroscience, 11(4), 398-403.

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. Mind & Society, 11(1), 3-13.

Tversky, A., & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. Science, 185, 1124–1131.

Winkler, R. L. (2015) The Importance of Communicating Uncertainties in Forecasts: Overestimating the Risks from Winter Storm Juno. Risk Analysis, 35: 349–353. doi: 10.1111/risa.12398.

Zhou, J., Sun, J., Chen, F., Wang, Y., Taib, R., Khawaji, A., & Li, Z. (2015). Measurable decision making with GSR and pupillary analysis for intelligent user interface. *ACM Transactions on Computer-Human Interaction (ToCHI)*, *21*(6), 33.