School of Computing Sciences, Postgraduate Research Day 2016: Book of Abstracts

5th October 2016

Table of Contents

16. Nick Matthews*, Stephen Laycock, Steven Hayward, GPU based regular grid construction for high quality deformable protein rendering.

17. *Seth Nixon*, Graham D. Finlayson*, Sudoku Texture Classification.

18. *Claudia Paicu*, Irina Mohorianu, Matthew Stocks, Tamas Dalmay, Vincent Moulton, Simon Moxon*, miRCat2: fast and accurate prediction of plant and animal microRNAs.

19. *Mohammad Ryalat*, Mark Fisher, Stephen Laycock*, An Investigation of Algorithms for the Automated Construction of 3D-Printed Immobilisation Masks for Use in Radiotherapy Treatment of Head-and-Neck Cancers.

20. *Awat Saeed*, Gavin C. Cawley, Tony Bagnall*, Differential weighting of Labelled and Unlabelled Patterns.

21. *Guillaume Scholz*, Katharina Hube*r, Representing orthology relations beyond trees.

22. *Ausdang Thangthai*, Ben Milner, Sarah Taylor,* Visual speech synthesis using dynamic visemes, contextual features and DNNs

23. *Kwanchiva Thangthai*, Richard Harvey*, Recent progress in Machine Lip-reading.

24. *Danny Websdale*, Thomas Le Cornu, Ben Milner*, Objective measures for predicting the intelligibility of spectrally smoothed speech with artificial excitation.

25. *Adam Ziolkowski*, Joost Noppen*, Enabling Continuous Testing for Software Development in the Chemical Industry.

Speaker in the Prof Andrew Bangham seminar series:

Information architecture for the genomics era

Dr William Spooner,
Eagle Genomics,
Wellcome Genome Campus, Hinxton, Cambridge.

Genomics is one of the defining innovations of our generation. It promises transformational precision products from medicines to toothpaste to seeds. This potential is reflected by government and industry investments totalling $Bns/yr. The field has progressed enormously since the first genomes of the late 1990s but there are still huge challenges to overcome.

The current bottleneck in genomics is bioinformatics. Some talk of "The $1000 Genome, the $100,000 analysis". We should not be surprised; adapting to data-driven world requires radical change. Similar "big data" problems have been faced (and solved) in other industries, but what principles can we learn from their successes? And in what respects is genomics exceptional? Insight into these questions can be gained through information architecture (IA), the structural design of shared information environments: It is not just how we design the pieces but how we put them together that matters.

For the past eight years Eagle Genomics has been pioneering IA for the genomics era. Our objective is to make data FAIR (findable, accessible, interoperable and reusable). Following the adage that "you can't manage what you can't measure" I present our recent work on the the systematic measurement of data value, and how this can form an integral component of the information workflow. Conceptual modelling is one thing, but does it work in practice? I will make the case using specific examples from biomarker discovery and microbiomics to demonstrate that yes, it really does!

Online Usability Studies
Evaluation and Comparison of Online Usability Studies in Different Environment Settings

Abeer Alharbi and Pam Mayhew

HCI and Usability Engineering
{Abeer.Alharbi, P.Mayhew}@uea.ac.uk

Online usability study has emerged that can be attempted by a large, varied pool of users anywhere as long as an Internet connection is existed. Would its usage help to give comprehensive insight of the whole user experience? That is especially interesting if the user operates remotely, as we are unaware of what the users might experience while performing the test. Using online usability study; this research aims to investigate whether the online usability study performance outcomes the same in a lab environment (quite environment and distractions and apparatuses usage are controlled) and a natural environment (open to distractions and any device can be used)? In other words, does the environment have an impact on the outcomes of an online usability study? If the online usability study performance outcomes are similar in both environments, then, can an online usability study be an alternative method to traditional usability testing?
An experimental study of an online usability study in two environment settings (lab and user's natural environment). In addition, the traditional usability testing in lab was included to serve as a benchmark. A total of 144 participated, 48 for each environment setting. All of them studying in UEA's different schools. UsabilityTesting tool was used to collect the data of the online usability study which automatically collect testing and participants' data. Direct observation, Retrospective Think Aloud Protocol and Camtasia Recorder 8 were used to collect the data of the traditional usability testing. Participants have been asked to perform three searching tasks on three distinct digital library websites and a control task with Amazon.co.uk website.

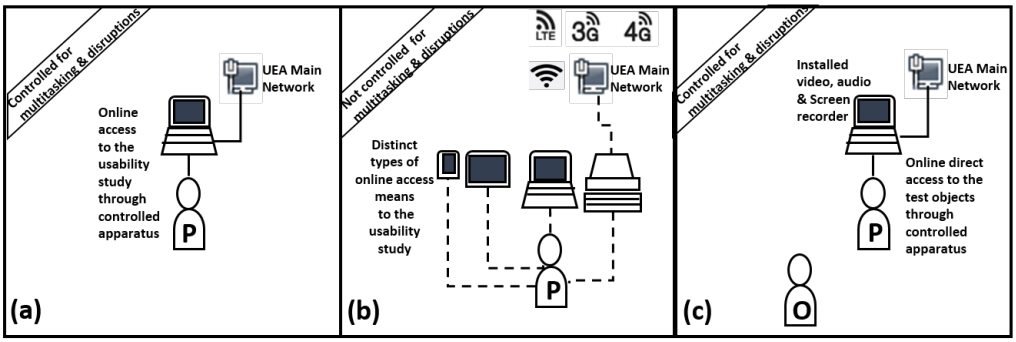

Figure 1: Type of testing environment settings (P: Participant; O: Observer). (a) Setting for applying online usability study in the lab, (b) Models of accessing the online usability study in the user's natural environment, (c) Setting for applying traditional usability testing in the lab

The study showed that no differences were found in the all of the measurements on the task level (time on task, page views, subjective evaluation of the website' usability, successful completions, the number of usability issues and their severity ratings). Yet, significant differences were found in time on test which is mainly caused by the large difference in the time on questions which was found to be influenced by multiple factors as interruptions, device type, internet connection and English language level. Participants' emotions were similar when performing online usability study, yet, the intensity of the positive emotions was decreased with the traditional usability testing where the test observer was present.
The study showed that online usability study performed similarly to the traditional usability testing. Based on this research findings, we can say that online usability study could be used as alternative to the traditional usability testing only if carful and rigid design was devoted to its design and if the online usability tools provide more technological support to the usability researchers and/or practitioner requirements.

Analytics of Information Management in Higher Education

Zahyah Alharbi, James Cornford, Garrick Fincham, and Beatriz De La Iglesia

Machine Learning and Statistics Laboratory
{z.alharbi,j.cornford, g.fincham, b.iglesia}@uea.ac.uk

There is currently an increasing interest on educational data mining. Higher Education institutions require an improvement of their educational quality to be more competitive; therefore the application of data mining on this setting is becoming very interesting to both university administrators and researchers. On the other hand, recommender systems are widely utilised in various areas, mainly in e-commerce. Lately, they are also employed in learning tasks such as recommending appropriate modules, books, papers etc. to the learners (students).

In this research, we investigate the use of data mining techniques in an educational setting to highlight performance problems early on and propose remedial actions. We also investigate recommender systems in the educational setting. We propose a recommender system that may guide students towards better module choices to increase their chances of a good outcome, based on their performance and other similar students' performances. We compare different prediction models in the context of recommender systems. We use real data from the data warehouse of a specific University. We validate our results by utilising data relating to students with different characteristic from different schools. We use additional data that could be available to improve the models, for example on engagement and attendance. We will also investigate how to make the recommender system acceptable to students, and how to utilise the available information to improve students' outcomes. We will achieve this through preparing a questionnaire and several interviews. Our research's end results will enable us to provide recommendation about the quality of the used data to improve the University data warehouse, about the technical aspects of building a recommender system and about the management aspects of deploying such system.

In-Depth Evaluation of Software Project Human Resource Allocation Optimization Approaches

Sultan M. Al Khatib, Joost Noppen, Pam Mayhew

Machine Learning
{s.al-khatib,j.noppen,p.mayhew}@uea.ac.uk

In software projects, managers are faced to perform the human resource allocation by staffing the available resources according to the estimated effort and skills required of project tasks considering the resource productivity and expertise, and then scheduling these tasks to minimize the project time. This problem of Staffing and Scheduling a Software Project (SSSP) in order to minimize the project time has been addressed by different approaches as in [1, 2, 3, 4, 5, 6, 7]. However, these approaches tend to vary in the parameters they consider such as resource skills, productivity measures, and optimization techniques for the same cost function. This variation between the SSSP approaches makes it difficult for the software manager to select the proper approach for the problem without knowing their accuracy, performance, and applicability of each. Only two studies in [8, 9] compare the SSSP approaches by either a comprehensive survey, or systematic literature review. Nonetheless, these studies are limited to the conceptual aspects they consider and not to the actual details of results that can be obtained by each based on their implementation and a reference dataset for a unified basis. We have introduced for this problem a systematic approach to compare the SSSP approaches. With our establishment of preliminary benchmark dataset, classification of the approaches, and performance measures, a comparison between five well known approaches is performed [3, 4, 5, 6, 7]. The results show that the approaches belong to the same class can differ by using the measurements proposed and that our approach is capable to identify these differences. The intended work is to include more approaches in this comparison and expanding the benchmark dataset. Moreover, an empirical evaluation for these approaches is planned to compare them with the industrial settings and validating them according to the current organizational practice. The results from this comparison will also be used to evaluate and to establish the relevance and accuracy of our method in real-world scenarios. The project afterwords can also focuses on improving on state of the art models for resource allocation by explicitly considering a prominent omission elements such as Agile and team formation of cross-functional expertise. This will be combined with state of the art optimization algorithms, such as genetic algorithms, to contain complexity and ensure the usability of our approach.

References

[1] H.-T. Tsai, H. Moskowitz, and L.-H. Lee, "Human resource selection for software development projects using taguchis parameter design," *European Journal of Operational Research*, vol. 151, no. 1, pp. 167 – 180, 2003.

[2] M. Di Penta, M. Harman, and G. Antoniol, "The use of search-based optimization techniques to schedule and staff software projects: an approach and an empirical study," *Software: Practice and Experience*, vol. 41, no. 5, pp. 495–519, 2011.

[3] C. K. Chang, M. J. Christensen, and T. Zhang, "Genetic algorithms for project management," *Annals of Software Engineering*, vol. 11, no. 1, pp. 107–139, 2001.

[4] E. Alba and J. F. Chicano, "Software project management with {GAs}," *Information Sciences*, vol. 177, no. 11, pp. 2380 – 2401, 2007.

[5] D. Kang, J. Jung, and D.-H. Bae, "Constraint-based human resource allocation in software projects," *Software: Practice and Experience*, vol. 41, no. 5, pp. 551–577, 2011.

[6] G. Antoniol, M. D. Penta, and M. Harman, "A robust search-based approach to project management in the presence of abandonment, rework, error and uncertainty," in *Software Metrics, 2004. Proceedings. 10th International Symposium on*, Sept 2004, pp. 172–183.

[7] G. Antoniol, M. Di Penta, and M. Harman, "Search-based techniques applied to optimization of project planning for a massive maintenance project," in *21st IEEE International Conference on Software Maintenance (ICSM'05)*, Sept 2005, pp. 240–249.

[8] F. Ferrucci, M. Harman, and F. Sarro, *Search-Based Software Project Management*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 373–399.

[9] D. C. C. Peixoto, G. R. Mateus, and R. F. Resende, "The issues of solving staffing and scheduling problems in software development projects," in *Computer Software and Applications Conference (COMPSAC), 2014 IEEE 38th Annual*, July 2014, pp. 1–10.

[10] Y.-K. Kwok and I. Ahmad, "Benchmarking and comparison of the task graph scheduling algorithms," *Journal of Parallel and Distributed Computing*, vol. 59, no. 3, pp. 381 – 422, 1999.