IITB-Monash Research Academy

An Indian-Australian research partnership

Indian Institute of Technology Bombay | MONASH University

CSIRO

Project Title:	**Deep User Models for Visual Analytics**
Project Number	**ID00729**

Monash Main Supervisor (Name, Email Id, Phone)	Tim Dwyer tim.dwyer@monash.edu	*Full name, Email*
Monash Co-supervisor(s) (Name, Email Id, Phone)		
Monash Head of Dept/Centre (Name,Email)	Kim Marriott kim.Marriott@infotech.monash.edu.au	*Full name, email*
Monash Department:	Monash Adaptive Visualisation Lab, Faculty of Information Technology	
Monash ADRT (Name,Email)		*Full name, email*
IITB Main Supervisor (Name, Email Id, Phone)	Venkatesh Rajamanickam venkatra@iitb.ac.in	*Full name, Email*
IITB Co-supervisor(s) (Name, Email Id, Phone)		*Full name, Email*
IITB Head of Dept (Name, Email, Phone)	G. V. Sreekumar gvsree@iitb.ac.in	*Full name, email*
IITB Department:	Industrial Design Centre, School of Design	
Data61 Supervisors	Ulrich Engelke, Cecile Paris, Matt Bolger ulrich.engelke@data61.csiro.au	

Research Clusters:

Highlight which of the Academy's CLUSTERS this project will address?
*(Please nominate JUST **one.** For more information, see www.iitbmonash.org)*

1	*Material Science/Engineering (including Nano, Metallurgy)*
2	*Energy, Green Chem, Chemistry, Catalysis, Reaction Eng*
3	*Math, CFD, Modelling, Manufacturing*
4	***CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control***
5	*Earth Sciences and Civil Engineering (Geo, Water, Climate)*
6	*Bio, Stem Cells, Bio Chem, Pharma, Food*
7	*Semi-Conductors, Optics, Photonics, Networks, Telecomm, Power Eng*
8	*HSS, Design, Management*

Research Themes:

Highlight which of the Academy's Theme(s) this project will address?
(Feel free to nominate more than one. For more information, see www.iitbmonash.org)

1	***Advanced computational engineering, simulation and manufacture***
2	*Infrastructure Engineering*
3	*Clean Energy*
4	*Water*
5	*Nanotechnology*
6	*Biotechnology and Stem Cell Research*

The research problem

Deriving insight from data is instrumental to any industry and government organisation for informed decision making and accelerated business agendas. The path from Data-to-Insight (D2I), however, is typically cumbersome involving tedious manual handling of data, complex analytics and often misleading visualisation techniques. Visual analytics alleviates this issue by integrating computational and human efforts, allowing for effective data exploration through interactive user interfaces. The challenge remains to optimise and automate the visual analytics process through specification of the D2I ecosystem, incl. data characteristics, system properties, analytics processes, and visualisation techniques, amongst other factors. Elaborate and widely accepted bodies of work exist that provide said specifications, such as the 'Grammar of Graphics' by L. Wilkinson. However, such specifications are typically used in isolation disregarding interactions within the ecosystem and all too often leaving out an instrumental factor in the visual analytics process: the human analyst.

Developing Deep User Models (DUMs) and integrating these into the visual analytics process is at the heart of this proposal. The ultimate goal is to optimise user performance and experience for a breadth of end users ranging from scientists, who need to gain deep insight into data, to executive decision makers, who need to be provided with only high level summary statistics. For this purpose, user models should include explicit knowledge about the user (e.g. expertise, skills, domain knowledge) as well as implicit knowledge (e.g. preference for certain information or visualisation types). The former can be specified in user profiles whereas the latter can be learned from the interaction of the user with the system.

In this project, the student needs to develop DUMs based on user profiles and learned interactions of users with the system. A large scale survey study with domain scientists and data analysts (e.g. from CSIRO, Australia/India, or globally) needs to be performed to define user profiles and develop a typology of data analysts. The user models are then to be integrated into the visual analytics process and iteratively updated based on interaction. The performance of the visual analytics process with and without the inclusion of the user models needs to be evaluated based on quantitative measures (e.g. completion time, task accuracy, physiological measures) and overall user experience. Existing analytics workflow tools, such as CSIRO's Workspace, may be used as a development framework and in turn may benefit from the inclusion of the DUMs.

Project aims

The aims of the project are threefold:

Firstly, the projects aims to gain a deep understanding of how domain scientists and data analysts arrive at a particular insight during the visual analytics process. This is facilitated through a large scale survey as well as interaction logging during the visual analytics process that lead to Deep User Models (DUMs).

Secondly, the projects aims to optimise the visual analytics process by tightly integrating the DUMs. This should lead to a more effective path from data to insight as the system can provide recommendations and guide analysts as they progress through the analytics process.

Finally, the projects aims to establish meaningful, quantitative measures of evaluating the performance of a visual analytics pipeline leading to specific insight. This should involve integrating existing measures (e.g. completion time, task accuracy) with newly developed measures, for instance, based on interaction logs or physiological measurements.

Expected outcomes

- An extensive taxonomy of data analysts based on a large scale survey.
- Deep User Models developed from the taxonomy and from learned interactions.
- A visual analytics system that integrates the Deep User Models.
- A set of evaluation methods that can quantify the performance of a visual analytics process.

How will the project address the Goals of the above Themes?

The project is well aligned with the 'Advanced Computational Engineering, Simulation and Manufacture' theme and specifically with the Computer Graphics and Computational Science expertise required within that theme. The project supports the challenge of handling extensive amounts of data about a system or process, as stated on the theme's website, by creating models about the end users of such a data handling process. The project is further well aligned with the 'CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control' cluster as it aims to optimise the data analytics process through the user models.

Capabilities and Degrees Required

Degree:
Computer Science, Human Computer Interaction, Data Visualisation, or a related discipline

Capabilities:
Motivation and discipline to carry out autonomous research
Ability to work effectively as part of a multi-disciplinary, regionally dispersed research team
Knowledge of user experience and visualisation design
Advanced programming skills (Python, Javascript, C/C++, C#)
Experience working with visualisation libraries such as D3.js, three.js, WebGL, and WebVR

Potential Collaborators

IBM Research, USA

Select up to **(4)** keywords from the Academy's approved keyword list **(available at http://www.iitbmonash.org/becoming-a-research-supervisor/))** relating to this project to make it easier for the students to apply.

Data Science, optimisation, algorithms (6)
Natural Language Processing (29)