Impact case study (REF3b)

Institution: Robert Gordon University

Unit of Assessment: 11 Computer Science and Informatics

Title of case study: Data-driven Decision Support

1. Summary of the impact (indicative maximum 100 words)

Many organisations rely on increasingly large and complex datasets to inform operational decision-making. To assist decision-makers when decisions are data-driven, computational tools are needed that present reliable summary information and suggest options allied to the key objectives of decision-making. Research at RGU has developed novel learning and optimisation algorithms driven by multifactorial data and implemented this in commercial decision-support software. The research has had economic impact by providing products to be sold: drilling rig selection tool (ODS-Petrodata Ltd.) and subsea hydraulics diagnostic tool (Viper Subsea Ltd.). Further economic impact comes from operations management software developed for British Telecom.

2. Underpinning research (indicative maximum 500 words)

Our research expertise in machine learning for complex decision support dates from 1999 when we first applied evolutionary algorithms to optimise chemotherapy treatments using data-validated tumour simulations [R1]. As decision-making has become increasingly data-driven, we have researched approaches that automatically detect interactions between decision variables in multifactorial datasets that are pertinent to key outcome variables. Such relationships when discovered can be presented to decision-makers directly and/or combined with methods that automatically suggest one or more decisions for consideration. In much of our work this sort of relationship may only be implicitly derived from running a simulation based on the decision variables or by analysing large historical datasets containing both decision and outcome variables. An example is the relationship between the selection, intensity and phasing of anti-cancer drugs and the outcomes in terms of tumour reduction, patient quality of life and the severity of toxic side effects. [R2] was the first work to apply multi-objective naturally-inspired algorithms combined with tumour response simulations and clinical trials data to support clinicians in developing improved cancer chemotherapies. All of these approaches are computationally intensive and complexity grows exponentially with dataset size, number of decision variables and number of choices per decision. This is problematic because real-world decision-support systems must be responsive to decision-makers and be implemented in a typical industrial IT environment.

Our research has therefore sought computationally cheap methods of data-driven decision-support, developing state-of-the-art learning and multi-objective optimisation algorithms that work with simulations and datasets and evaluating them against comparable techniques on theoretical and on real world problems. There are two main approaches in our work.

search efficiency: we build cheap but inaccurate data models during optimisation to quickly focus the search on feasible, near-optimal regions of the search space. More expensive and accurate computation can then be focussed on a small number of highly promising solutions. In 2007 we introduced the Chain-Model Genetic Algorithm (ChainGA) for Bayesian Network Structure Learning [R3]. Significant speed-up was achieved using simple chain models to quickly evaluate candidate solutions. Comparisons with a classical algorithm (K2GA) showed that the benefits of ChainGA were highly problem-dependent. Further research has reinforced this finding and identified how search space and algorithm features interact to affect search effciency [R4].

search accuracy: We use sophisticated learning to greatly reduce computational effort through needing fewer solution evaluations. This is particularly appropriate where the costs of model building are dwarfed in comparison to the costs of evaluating solutions. The DEUM algorithm builds a Markov network model of the relationship between solutions and fitness during an optimisation. In 2007, we compared DEUM against other leading EDAs [R5] showing that it significant reduced the number of function evaluations needed for optimisation at the expense of computational cost. More recent work (2008 onwards) has focussed on analysing the number of evaluations needed in relation to known or discovered problem structure. Precise estimates of the

number of evaluations required have now been shown to hold valid for a wide range of theoretical and applied problems [R6]. Therefore robust estimates can be made of the computational cost trade-offs between solution evaluation and model-building for new applications based on the nature and number of the variables involved and simulation run-times.

Key researchers:

John McCall (Lecturer > Senior Lecturer > Reader > Professor (1992 – date)

Andrei Petrovski Lecturer > Senior Lecturer > Reader (1998 - date)

Siddhartha Shakya PhD Student > Research Fellow (2003-7) > Research Manager, British Telecom Innovate and Design (2007 – date)

---

3. References to the research (indicative maximum of six references)

* indicates key reference.

[R1] J. McCall, A. Petrovski (1999), *A Decision Support System for Chemotherapy using Genetic Algorithms*. In M. Mouhammadian (Ed.) *Computational Intelligence for Modelling, Control and Automation*, pp 65-70, IOS Press, 1999. [32 Google Scholar citations (10 self)]

* [R2] Andrei Petrovski and John McCall (2001), *Multi-objective Optimisation of Cancer Chemotherapy Using Evolutionary Algorithms*, in E.Zitzler et al. (Eds.): *Evolutionary Multi-Criterion Optimization*, Lecture Notes in Computer Science **1993**, pp531-545, Springer-Verlag ISBN: 3-540-41745-1, ISSN:0302-9743. [33 Google Scholar citations (11 self)]

[R3] Ratiba Kabli, Frank Herrmann, John McCall, A Chain-Model Genetic Algorithm for Bayesian Network Structure Learning, in GECCO 2007 Volume II pp1264-1271. [22 Google Scholar citations (12 self)]

[R4] Wu, Y., McCall, J., & Corne, D. (2010). Two novel Ant Colony Optimization approaches for Bayesian network structure learning. In Evolutionary Computation (CEC), 2010 IEEE Congress on (pp. 1-7). IEEE. [15 Google Scholar citations (7 self)]

* [R5] Siddhartha Shakya, John McCall (2007), *Optimization by Estimation of Distribution with DEUM Framework Based on Markov Random Fields* Int. Jnl of Automation and Computing **04 (3)**, 262-272 DOI: 10.1007/s11633-007-0262-6 [53 Google Scholar citations (16 self)]

* [R6] Alexander E.I. Brownlee, John A.W. McCall, Siddartha K. Shakya and Qingfu Zhang (2009), Structure Learning and Optimisation in a Markov-network based Estimation of Distribution Algorithm, in Proceedings of the Eleventh IEEE on Congress on Evolutionary Computation, pp447-454, 2009. [27 Google citations (10 self)]

---

4. Details of the impact (indicative maximum 750 words)

Pathway to Impact

Decision support technologies have a complex pathway to impact. We operate in areas where complex decision-making controls the use of large assets or the deployment of workforces. Here operational expenditure is large and so are the consequences of inefficient decision-making. Therefore the process of adopting innovation in decision-making policy is multi-staged and conservative. Some large companies innovate in-house – an example is British Telecom. Others, such as oil and gas majors, outsource innovation to technology support companies, which are often SMEs. Examples of such support companies here are Viper Subsea and ODS Petrodata.

Decision-support technologies are therefore developed through stages of "technology readiness" before ultimate internal adoption or external purchase by a large company. There is therefore a value chain of technology adoption. As an academic research group, we operate at the start of the value chain by translating advanced knowledge in optimisation and learning into software components that address a decision-making problem identified by a company as an area for improvement. We essentially operate as an outsourced part of these companies' R&D functions and add value by proving concepts, quantifying potential efficiencies and encapsulating specialist know-how in software components around which decision-support products can be developed and sold. Our main economic impact therefore is in creating this value. It is measured by proven

efficiencies in decision-making processes and the potential sales value this creates for the companies, or, in the case of a large company, internal adoption for use in high-value decision-making. Typically we operate by embedding our research students with partner companies and often this leads to employment to support development through the next stages of the value chain. Therefore, a secondary economic impact is the transfer of high-skilled individuals into companies.

Reach

In this case study we explain how our research in data-driven decision support has impacted one large company, British Telecom, and two SMEs, ODS Petrodata and Viper Subsea. In each case, the interaction has resulted in value-add in terms of decision-support software and the long-term employment of research students by the company. We discuss specific aspects of the pathway and significance for each company in turn.

Operations Management Decision Support at British Telecom

Context

Workforce Dynamics Simulator (WDS) is a dynamic business simulation environment, developed at British Telecom, which enables large scale simulation of workforces and notable resources. It uses historical or generated data to investigate the execution of work plans. WDS takes detailed input data on geographical locations, tasks to be completed and engineers available, typically over a 90-day planning period and involving around 25,000 workers across the UK. The system simulates a variety of scenarios and the output reports are used to support decision making in the business. A key problem is that WDS is sensitive to choice of input parameters. These must be tuned to ensure that the predictions in the output report are as accurate as possible.

Pathway

In 2010-11, BT funded a project at RGU supervised by **McCall**, to apply metaheuristic optimization algorithms to carry out the tuning of WDS. WDS simulations are expensive and accuracy is important so **search accuracy** was a target here. A variety of different algorithms were evaluated, including the DEUM algorithm. The decision to investigate the use of DEUM in this case was motivated by the high computational cost incurred with each simulation. We packaged a suite of algorithms, including DEUM, as a software component called the WDS Tuning Tool (WDSTT).

Impact and Significance

WDSTT has been adopted by BT, replacing a parameter sweep of tuning parameters for WDS. WDSTT adds value in up to 5% more accurate field workforce simulations as well as reduced tuning times (up to a factor of 10). Since the incorporation of WDSTT, the simulation software has been used repeatedly in BT to support management decision-making in assessing the effect of changes in different business scenarios and transformation initiatives including: the effect of new product introduction; the effect of special events on demand; and support for IT investment decisions [I1]. It is impossible to quantify financially the effect of more accurate simulation on the quality of decisions. However the decisions involved correspond to investment of multi-million pound amounts in the deployment and organisation of over 20,000 field workforce employees, so small percentage changes in accuracy translate into significant financial and operational value.

McCall's former research student (**Shakya**) is permanently employed as a research manager at British Telecom. While at BT, he has applied DEUM to a dynamic pricing problem [I2].

Subsea Hydraulic Control Diagnostics at Viper Subsea

Viper Subsea (VS) are an engineering company offering specialised consultancy support for hydraulic control systems for offshore oil and gas production. Offshore drilling installations consist of an interconnected set of pipes, pumps and valves used to collect hydrocarbons from different wells for onward processing. Pumps and valves are controlled hydraulically through hydraulic controls. Due to the nature of hydrocarbon flow, pipes pumps and valves can become clogged or stuck and so performance changes over time. Shutdowns in production cost millions of dollars per day and so it is important to correctly diagnose faults in the system when they occur, either to avoid or to minimise and to accurately target shutdowns. Viper Subsea are therefore interested in developing diagnostic software for sale to operators of offshore installations.

Pathway

VS and Technology Strategy Board (TSB) funded a two-year KTP project (2011 – 13), supervised by Petrovski and McCall, to develop algorithms for pro-active fault diagnosis using control feedback data from the real-time operation of subsea hydraulic systems. As the algorithms need to be accurate but also run on limited capacity machines in robust environments search efficiency and search accuracy are both key. A wide range of learning algorithms, including Bayesian Network learning [R3, R4], was evaluated on test problems, using test data supplied by a major operator. A diagnostic software component was packaged in a prototype product V-Sentinel, which can, within a few seconds, identify abnormalities in data streams from 14 sensors, characterising the level of hydrolic liquid in a tank, quality of insulation, and validating pressure sensors.

Impact and Significance

In a final report on the project, required by TSB, VS valued V-Sentinel as having a profit-generating capacity of £165K per annum [I3]. It is reasonable to value an asset at five times its annual profit-generating capacity, in this case £825K. VS are actively developing V-Sentinel and the research student is now permanently employed by the company for that purpose.

Drilling Rig Market Data Modelling at ODS Petrodata

Context

ODS-Petrodata Ltd. (ODSP) are an Aberdeen-based SME specialising in the provision of oil and gas market data. ODSP's market data and analyses are made available to the oil and gas market through a number of web products including RigPoint, which oil and gas majors and large financial investors use to inform their selection of offshore drilling rigs for particular jobs. The drilling performance of a given rig depends on a large number of factors relating to rig specification, drilling location and water depth Given that each rig-hire costs several hundreds of thousands of dollars per day, the potential benefits of reduced drilling time by a few days on average, realised across the hundreds of annual hires in the global market, run to $10Ms per annum. ODSP were therefore interested in adding value to their RigPoint product with a drilling rig selection tool, able to inform choice by predicting time to drilling depth for selected rigs.

Pathway

ODSP and TSB funded a three-year KTP project (2009 – 12), supervised by **McCall** and **Petrovski**, focussed on integrating and modelling two large datasets containing over one hundred and fifty data fields and fifty thousand records of drilling rig and well data. The size of the datasets and consequently modelling complexity necessitates **search efficiency**. Chain-model algorithms [R3, R4] were used to produce Bayesian Network models relating the data factors. These models were incorporated in a software tool to predict the drilling performance of particular rigs given a range of over forty possible input factors that would be known to decision makers, including drilling location and depth. The software is capable of inferring factors that are not given.

Impact and Significance

In the final report on the project, required by TSB, ODSP valued the software component as having a profit-generating capacity of £1.35M per annum [I4], so an estimated value of £6.75M. As the project was nearing completion, ODSP was acquired by a large US data company IHS Ltd. IHS committed to continuing development of the rig efficiency software [I4] and the research student was taken on as a permanent employee.

---

5. Sources to corroborate the impact (indicative maximum of 10 references)

[I1]   Letter of corroboration from Head of Resource Management Technologies Research, British Telecom Innovate and Design.

[I2]   Shakya, Siddhartha, Fernando Oliveira, and Gilbert Owusu. "Analysing the effect of demand uncertainty in dynamic pricing with eas." *Research and Development in Intelligent Systems XXV*. Springer London, 2009. 77-90.

[I3]   Partners Final Report Form: Knowledge Transfer Partnership KTP008580.

[I4]   Partners Final Report Form: Knowledge Transfer Partnership KTP006922.