International Journal of Information Management 36 (2016) 607–617

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Model and experimental development for Business Data Science

Russell Newman [a,*], Victor Chang [b], Robert John Walters [a], Gary Brian Wills [a]

[a] Electronics and Computer Science, University of Southampton, Southampton, UK
[b] Xi'an Jiaotong Liverpool University, Suzhou, China

ARTICLE INFO

Article history:
Received 2 April 2016
Received in revised form 2 April 2016
Accepted 2 April 2016
Available online 8 April 2016

Keywords:
Business Data Science (BDS)
Modelling and experimental techniques for BDS
Future directions and contributions for Data Science

ABSTRACT

While Data Science has become increasingly significant for business strategies, operations, performance, efficiency and prediction, there is little work on this to provide a detailed guideline. We have proposed a Business Data Science (BDS) model that focuses on the model and experimental development that allows different types of functions, processes and roles to work together collaboratively for efficiency and performance improvements. Details with examples have been illustrated to show that BDS model can be a robust model. Future directions have been discussed to ensure that business intelligence, security, analytics and research contributions to BDS can be achieved.

1. Introduction

Data Science is an interdisciplinary area to enable experts in different domains to study and work together (Borrego & Newswander, 2010; Provost & Fawcett, 2013a; Ericsson, 2014). Outputs from all kinds of work can generate data in different types of formats. It has become apparently obvious that the processing, analysis and presentation of data outputs will be important to a growing number of sectors involved (Agresti & Kateri, 2011). The main reason why Data Science makes attractive to businesses is: Data Science is a study of the data that has involved processing, analysis, interpretation and making sense of the data (Han, Kamber, & Pei, 2011; Gelman, Carlin, Stern, & Rubin, 2014). Businesses can understand their problems, their business performance (daily, weekly, monthly and yearly) and forecast of their business performance within a matter of minutes at any time (Provost & Fawcett, 2013a).

The role of Data Science has become increasingly important for businesses as follows. Firstly, Data Science allows businesses to collect and analyse data about their business operations, strategies and overall performance (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). Secondly, business can improve on their services, operations, strategies and business performance based on the outputs of analysis (Nath, Nachiappan, & Ramanathan, 2010). Thirdly, businesses can improve the quality of their predictive modelling, so that decision-makers can plan for suitable strategies for their companies (Dhar, 2013). There are three major benefits of doing so, however, the ways to execute Business Data Science (BDS) are not established as yet since existing literature does not have a conclusive guidelines or a summary of best practice approach. Although there are many organisations that have become interested in Data Science, they do not know how to operate and manage Data Science (McAfee et al., 2012). This has motivated us to present our case of BDS, particularly in the way that organisations can adopt. Furthermore, a structured guideline is useful for development of any projects and services.

In order to demonstrate how to be effective for businesses, our research is focused on development of relevant modelling and simulation techniques, to provide organisations a bridge and a smooth transition to the adoption of Business Data Science (BDS). The fundaments of these techniques are then used to construct a BDS model as explained throughout the paper. To ensure a BDS model can work effectively with business activities, modelling and simulation techniques are required to be investigated to ensure business models, functions, processes with different roles of people involved can be resilient and robust. The structure of this paper is as follows. Section 2 explains the definitions, scopes, components, functions and overall approach towards modelling and simulation techniques for BDS. Section 3 presents the experimental design for BDS model that blends business intelligence, investigation to economic bubbles and other related areas. Section 4 presents four topics of discussions and Section 5 sums up this paper with the future work described.

* Corresponding author.
E-mail addresses: rn2@ecs.soton.ac.uk (R. Newman), ic.victor.chang@gmail.com (V. Chang).

http://dx.doi.org/10.1016/j.ijinfomgt.2016.04.004
0268-4012/© 2016 Published by Elsevier Ltd.

2. Modelling and simulation techniques for business data science

Modelling and simulations techniques are useful for business to stay competitive, efficient and collaborative. Understanding the terminologies, including what each term means and how each terminology offers is also relevant for business growth and sustainability (Chang, 2015a). Their definitions are as follows.

Simulating a system enables analysis of various situations by modelling them, over time, within a computer program (Banks, 1998).

A model is a “representation of an event and/or things that is real (a case study) or contrived (a use case)”. A simulation is “a method for implementing a model over time” (Banks, 2009).

A simulation may be run multiple times, to investigate how differing conditions alter the outcome. The competency to manage and master business data science has become significant for organisations that adopt business intelligence and analytics approach (Chen, Chiang, & Storey, 2012).

The word “system” denotes what, from the real world, is being simulated. A system may be broken down into its composite elements (such as people, machines and resources).

The “system model” is the simulated representation of the real-world system. System models are designed and built in such a way that a computer can perform calculations upon them and effectively run a simulation (Banks, 2009). When interpreting a system to build a model, an important consideration is how the various elements are to interact and affect one another (Cellier & Greifeneder, 2013).

The actual *method* for designing the model depends upon which simulation technique is employed. Regardless of the technique employed, a model designer must determine the level and areas of detail for the model, known as *scope*.

2.1. Scope

Scoping is the process of deciding which components should be included in a simulation models, and at what level of detail, and which components should be left out, simplified or abstracted (Sokolowski, 2009).

Recording a real-world system into a quantified model means that some concessions and assumptions must often be made. For instance, a certain component of a real-world system could be implemented fully in a simulation, resulting in a theoretically accurate simulation, at the cost of including many elements to represent the system. Alternatively, the same component may be modelled using a simpler implementation that is perhaps abstracted or makes some assumptions, without compromising the accuracy of the rest of the model (Cellier & Greifeneder, 2013).

This example demonstrates how detail or depth of scope must be decided. A model designer must choose exactly how much detail should be expressed in a model; greater detail may lead to a more accurate model, but at the same time create potentially unnecessary work in situations where a less detailed model would be sufficient.

The breadth of the model must also be scoped. A system model may contain modelled representations of many external entities that influence the core elements of the simulation. Including more of these may increase accuracy, again at the expense of time and design complexity. Alternatively, excluding more of these may result in an adequate simulation model and a saving of time.

2.1.1. System Dynamics

System Dynamics is a method of quantifiably modelling and simulating complex systems. It was developed by Jay Forrester at the MIT Sloan Management School, which was founded to exploit a fusion of engineering tools and techniques with traditional management. It was initially developed as a means of identifying the factors that make up the success or failure of a corporation or group of people (Forrester, 1997). System Dynamics models were originally processed by hand, but the technique was later adapted to take advantage of computer processing.

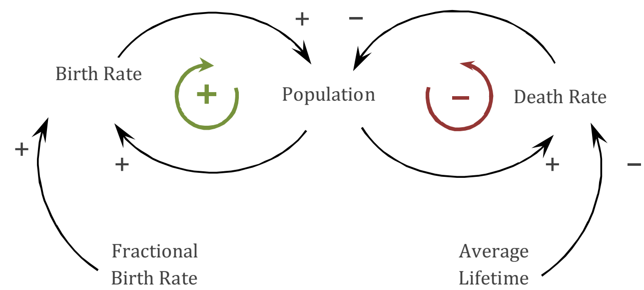

Fig. 1. Population System Dynamics model derived from Fontaine et al. (2009). (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

To build a System Dynamics simulation, a design progresses through two distinct stages: building causal loops and translating these to stocks and flows. It is possible to skip the first stage, but this would also exclude valuable analysis of the system which can lead to a higher quality simulation. Sterman's (2001) work on System Dynamics provides a succinct description of the methodology, which is used to inform the following sections.

2.1.2. Causal links and loops

At this first stage of design, a causal loop diagram is designed showing the various system variables and influences between each of them (Sterman, 2001). The causal loop diagram does not quantify any of the variables, but does denote whether variables positively or negatively affect each other. This is done through *feedback links*.

A feedback link is represented by an arrow. It signifies that the variable at its origin affects the variable at its destination in a positive *or* negative way relative to changes at the origin.

- A positive link means that as the item at the origin increases, the item at the destination may increase as well. If the origin were to decrease, the item at the destination would also decrease.
- A negative link means that as the item at the origin increases, the item at the destination may decrease. If the origin were to decrease, the item at the destination would increase.

Fig. 1 shows a simple causal loop diagram modelling the causes and effects between variables of birth rate, death rate and living population.

Using these constructs, loops are not an uncommon feature in designs. A loop emerges when feedback links connecting variables form a closed path. Loops may be categorised according to whether they result in a positive or negative affect after all components of the loop have been evaluated.

- A positive (reinforcing) loop is denoted with a+ (shown in green in Fig. 1). It emerges when, after evaluating all the links in the loop, there is an overall positive effect upon all variables within. This can lead to exponential growth of variables, if not mediated with other links.
- A negative (balancing) loop is denoted with a− (shown in red in Fig. 1). It emerges when, after evaluating all the links in the loop, the variables inversely affect each other.