Expert Systems With Applications 117 (2019) 330–344

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A data analytic benchmarking methodology for discovering common causal structures that describe context-diverse heterogeneous groups

Sergey Samoilenko[a,*], Kweku-Muata Osei-Bryson[b]

[a] Department of CIS/CS, Averett University, Danville, VA 23220, USA
[b] Department of Information Systems, Virginia Commonwealth University, Richmond, VA 23284, USA

ARTICLE INFO

Article history:
Received 3 March 2018
Revised 26 September 2018
Accepted 27 September 2018
Available online 27 September 2018

Keywords:
Benchmarking methodology
Information and communication technology capabilities
Causal structures
Data envelopment analysis
Market basket analysis
Decision tree induction
Cluster analysis
Association rules mining
Sub-Saharan Economies

ABSTRACT

Modern organizations typically regard information and communication technologies (ICTs) as one of the significant direct or indirect inputs for achieving operational excellence and competitive advantage. Since the concept of competitive advantage involves a relative comparison of the performance of organizational entities, then the concepts of organizational capabilities, context, and benchmarking are relevant. In this paper we present a new multi-method methodology for benchmarking that explicitly takes into consideration context-specific factors impacting the performance of organizational entities. This novel methodology allows for obtaining actionable information, in the form of non-obvious common causal structures, for improving the performance of the less efficient entities vis-à-vis their more efficient counterparts. The new methodology is state-of-the-art and is novel because it explicitly takes into consideration the context within which the organizational entities perform. Such "context awareness" allows for expanding the universe of discourse within which the process improvement initiatives are usually considered, thus allowing to consider the impact of external to the process factors on internal to the process mechanisms. This methodology involves the creative integration of several Information Systems (IS)' artifacts (i.e. multiple data mining methods) with Data Envelopment Analysis (DEA). We present an illustrative application of this methodology to an IS/ICT & Productivity research problem in the 'developing' countries context.

1. Introduction

It is commonly acknowledged that operational excellence is one of the sources of competitive advantage of modern enterprises. Fundamentally, the concept refers to achieving a high level of efficiency of conversion of inputs into outputs, where a higher level of efficiency implies, ceteris paribus, a greater degree of excellence. Modern organizations typically regard information and communication technologies (ICTs) as one of the significant direct or indirect inputs for achieving such operational excellence and competitive advantage. Since the concept of competitive advantage involves a relative comparison of the performance of organizational entities then the concepts of organizational capabilities and benchmarking (e.g., Gouveia, Dias, Antunes, Boucinha, & Inácio, 2015) are relevant. Ayabakan, Bardhan, and Zheng (2017) noted that with regards to the investigation of this pair of concepts: "*A dominant approach in IS research involves the use of survey instruments designed to elicit user responses on their perceptions about competencies and capabilities … A limitation of such perception-based approaches is that they represent a subjective measure of firm/organizational capabilities*"; these researchers therefore proposed an approach that involves the use of non-subjective data and the data envelopment analysis (DEA) method for doing benchmarking (e.g., Adler, Liebert, & Yazhemsky, 2013, LaPlante & Paradi, 2015) of the *Input-Output* conversion process. In this paper we also take a similar approach to benchmarking but are also interested in the context of the *Input-Output* conversion process, including those that involve ICTs as input(s). The motivating idea for this research project is that the process of benchmarking could possibly be enhanced by the discovery and application of non-obvious common causal structures that differentiate more efficient organizational entities from less efficient ones. This motivating idea triggered our intention to design an appropriate methodology artifact that involves the analysis of non-subjective data. This research project can be considered to fall within the realm of Information Systems (IS) research for at least the following reasons: (1) benchmarking research is an aspect of well-established IS/ICT & Productivity research stream (e.g. Hitt & Brynjolfsson, 1996; Ko & Osei-Bryson, 2004); (2) benchmarking research has appeared in leading IS journals (e.g. Ayabakan et al., 2017); (3) the proposed solution artifact involves the a creative in-

* Corresponding author.
E-mail addresses: ssamoilenko@averett.edu (S. Samoilenko), KMOsei@vcu.edu (K.-M. Osei-Bryson).

https://doi.org/10.1016/j.eswa.2018.09.054
0957-4174/

tegration of several IS artifacts (i.e. multiple data mining methods) with DEA; and (4) our illustrative example falls within the well-established IS/ICT & Productivity research stream.

The presentation of the project starts with our analysis of the concept of benchmarking, followed by the presentation of the research problem and associated research questions of the study. Then we offer to our reader an overview on supporting data analytic and data mining methods, with the methodology of the study presented next. The illustrative example demonstrates the application of the methodology and is followed by conclusion of the paper.

2. A conceptualization of the benchmarking problem

The term "benchmarking" is a commonly used one, and popular terms tend to be vulnerable to falling prey to the unfortunate assumption of the universality of their meaning. The concept of benchmarking is important to our inquiry; thus, we feel it is warranted if we spend a few sentences making sure that we clarify the chosen meaning of the term to our readers. Fundamentally and historically, benchmarking means *accurate application of a measure*, whatever the measure of interest could be (Merriam-Webster, 2017). Consequently, benchmarking is inherently a two-part process. Firstly, the presence of some sort of a standard measure, or a *target*, is established. And secondly, the process of emulation of that target is undertaken. For example, when we use a ruler to draw a 10-inch line on a piece of paper, we are benchmarking a chosen target (e.g., 10 inches) in the context of the piece of paper.

We can generalize benchmarking as *a process of emulation of the target in a new context*. This, immediately, brings up the problem of representation of the target, for in order to emulate a target we need to know the attributes that sufficiently describe (e.g., represent) the target. This is not an easy undertaking in the case of complex targets. It is one thing for a freshman to benchmark her academic performance by aiming to have 4.0 GPA (a trivial one-dimensional construct), and another thing for her to benchmark against the likes of Einstein (a non-trivial multi-dimensional construct). Similarly, it is easy for a competitor to benchmark the battery life of the cellular phone of the industry leader, and it is hard to benchmark the phone itself.

The whole concept of formal representation, which is relevant to benchmarking specifically and underlies the whole field of computing generally (e.g., if it can be formally represented, then it can be computed), is dependent on two factors. One is *objectivity* and the other is *scope*. The factor of objectivity of representation refers to having an objective (e.g., standard, agreed upon) scale for a given characteristic of interest, where an attribute "Sugar Content" could be objectively represented via "grams per kilogram" scale, and subjectively via the scale "perceived sweetness." The factor of objectivity, which is dealt with by finding or creating an appropriate measurement scale, is much easier to address than the factor of scope. In simple terms, scope deals with selecting what is "*in*" and what is "*out*", deciding on a set of attributes that adequately model (e.g., describe, represent) the target. The complexity of the decision is directly related to the complexity of the target; consequently, it is easier to benchmark a body mass index (BMI) than a luxury car, or a successful firm.

This issue of scope is not a trivial one because from a complex systems perspective we do not have a philosophical basis for making the decision regarding what the adequate description of the system itself would be. While a component-based description seems to be good and easy beginning, things get increasingly complicated once we start considering the non-linearity of the relationships between the components, various dependencies, and emergent properties. Furthermore, in the case of complex adaptive systems (e.g., person, organization, or economy) we cannot completely "abstract away" the system's environment, which further complicates the undertaking of a nice and neat scoping of the target. Under such circumstances, *the scope of the target of benchmarking is not given, but is a result of an active discovery*, and it is in this area that our paper aims to contribute.

While the concept of benchmarking can be applied to a great variety of contexts, in this paper we apply it to *economic units*, which we define as a set of entities (e.g., firm, economy) that: a) transform a set of expense-associated inputs into a set of revenue-associated outputs; and b) aim to minimize expenses and maximize revenues.

Any viable economic entity aims to ensure its survival by means of adopting and maintaining a valid business model. One of the purposes of a business model is to ensure that the stream of revenues is greater than the stream of expenses. While a valid business models assures an operational-level day-to-day viability of the entity, it does not a guarantee of a long-term survival, for an entity could be doomed due to failed strategic- and tactical-level initiatives, such as poorly chosen and implemented strategy, erroneous vision, or misguided business goals.

In order for a business model to bring the intended results it must be implemented- this is done via business processes. For example, a business model "sale of product to customers" in the context of a bakery can be implemented by means of the "acquire inventory", "bake products", and "sell baked goods" processes. It is an effective and efficient execution of business processes that constitutes the operational excellence of firms, and it is not surprising, therefore, that the less successful firms often aim to improve their operational performance by means of benchmarking of their business processes with those of their more successful counterparts.

Our focus in this research is on benchmarking of business processes within the context of economic units, and, conceptually, the problem that we are trying to address is associated with the necessarily different contexts of the target and the destination of the benchmarking. Let us consider two bakeries – one being a target of benchmarking (e.g., highly efficient) and another one being a destination of benchmarking (e.g., less efficient). Clearly, the process of baking is important and is easy to define based on the "*ingredients → baked products*" model. However, the same process takes place within different contexts- two bakeries may have different number of employees with different number of years of experience, they may have different equipment, and they may have different environmental conditions. Thus, the context of the process of baking is also important, but is difficult to define, for there is no common context model for the target and the destination.

Conceptually, the question can be expressed as:

> *How can we scope the target of benchmarking so that it takes into consideration contextual factors that are relevant and applicable to the destination of the benchmarking?*

3. Research problem and research questions of the study

Fundamentally, any business process can be seen as a process of conversion of means into ends, where the primary goal is to minimize the cost of means and maximize the value of ends via increasing efficiency and effectiveness of mechanism of transformation. If we consider a concept of a business process from a structural perspective, we can identify three distinctive parts. First, there is a set of inputs, second, there is a set of outputs, and third, there is a mechanism of transformation of inputs into outputs. This structural decomposition of a business process is important because it allows identifying a component, or components, that are relevant to the process of benchmarking.

A simplified model of conversion of inputs into outputs will take into consideration only those inputs that are necessary for