Proceedings of the 36th International Computers and Industrial Engineering Conference, C&IE 2006, June 2006, Taipei, Taiwan, pp.1-8.

The Data-Inspired Innovation Model

Andrew Kusiak and Chun-Yu Tang
Intelligent Systems Laboratory, MIE
3131 Seamans Center
The University of Iowa
Iowa City, IA 52242 – 1527
USA
andrew-kusiak@uiowa.edu

Abstract

To date numerous results on innovation research have been published, however, the literature on innovation does not offer conclusive findings and guidelines for practice. The research and practice of innovation are fragmented and centered on specific cases. The literature does not present a unified theory of innovation. Innovation in any domain could be enhanced by principles and insights from other disciplines. However, the process of identifying the linkages between the diverse disciplines and the application domain is not well understood.

The innovation process and conditions triggering innovation set the stage for economic progress. The paper discusses innovation as an outcome of analysis of data performed over the entire product-life cycle. The concept of data-driven innovation is described. While some of the data is collected in routine practice, other valued pieces of information are cultivated according to the practice of data farming.

Keywords: *Innovation process, data mining, knowledge discovery*.

1. Introduction

Innovation is defined as the activity of people and organizations to change themselves and the environment (Piana 2003). The scope of innovation varies from products and processes to organizations or even societies. The nature of innovation is user dependent, e.g., a product innovation for a designer can be a process innovation for a producer.
Nevertheless, innovation will only be successful if the customers perceive its value. The customers of today are better informed than ever before. The advent of the internet has expanded the scope and volume of information. The time scale of the interactions between customers and producers has been significantly reduced and it will undergo further reductions. Understanding customer's requirements and meeting their needs is no longer an option, but the way to succeed in the marketplace. While traditional market research techniques have been proven valuable in obtaining customer requirements, the need for innovation prompts for a more efficient and comprehensive process of information acquisition and analysis. Those who will make a better use of the information will gain an advantage.

The study of innovation – the development of new knowledge and artifacts – is of interest to engineering, business, social and behavioral sciences, and spans across sociology, history, philosophy, economics, psychology, and political science (Troyer 2005). Researchers from different areas have examined how innovation is formed, progressed, and is disseminated (Parthasarthy and Hammond 2002, Gopalakrishnan and Damanpour 1997, Klepper 1996).

Innovation in any domain can be enhanced by principles and insights from other disciplines. However, the process of identifying the linkages between different domains and the need for innovation science is apparent. Without a uniform platform for comparing different ideas, the knowledge of innovation will remain dispersed and fragmented. Innovation in product and process areas is of particular interest to manufacturing and service applications.

Typology of Innovation

Innovation studies have resulted in numerous typologies that are summarized in Tang and Kusiak (2006). Most of these typologies ignore the data aspect of innovation. This can be explained by the fact that the innovation studies have been performed before the information age.
Some of the most important dimensions of innovation from the data perspective are:

- New product vs existing product innovation. Innovation does not have to necessarily lead to the introduction of a new product (an invention), rather it may apply to the existing product. The data-driven approach to innovation discussed in this paper is intended for innovation of the existing products and processes; however, the implications may also lead to the creation of new products and processes. The decision in either direction usually has to do with the market acceptance and corporate strategy.
- Degree of innovation. Analogous to the level of system automation or intelligence, the degree of innovation may vary. One could argue that for any application at hand there is an optimal degree of innovation. It is somewhat unfortunate that metrics to adequately measure innovation ahead of its market validation are lacking.
- Scope of innovation. Innovation does not have to lead to a brand new product or process, irrespectively how it is desired. It may happen at a component, subassembly, or a process element level. For example, using a nano-tube in an existing product could revolutionize its design and make this product innovative.
  The scope of innovation may involve the following scenarios:

- Single product innovation
- Product family innovation
- Product portfolio innovation

It is conceivable that a company could concentrate on innovation of a single product (e.g., a pharmaceutical company developing on a new drug), a family of products, or the entire product offering. The same applies to processes or a product-process combination.

2. Exploring Innovation from Data

Innovation studies include a wide range of domains, different perspectives, and approaches. Some research findings were derived by studying narrowly defined targets thus making their generalization difficult. Other studies were designed to validate a general hypothesis using statistical methods or simulation tools. These findings that provided some understandings of how innovation was produced in the past may not be readily useful for innovation in the future. It is of paramount importance not only to know the determining factors of innovation but also to the interactions of them. The events triggering innovation and the process leading to the innovation are of particular interest.

Data-driven innovation is a new concept that may meet the needs of practitioners. Its underlying belief is that "new knowledge or valuable innovative ideas are embedded somewhere in the data." Since the collected data may document the environmental variables, the decision process and the limiting constraints, valuable insights are usually hidden in it. It is estimated that more than 75% of the new design initiatives use the previous design knowledge (Iyer *et al*. 2005). Past knowledge could be improved and applied in other areas.

Innovation has many faces, ranging from the evolutionary patterns in the product life-cycle to what the customer values the most. Any product that has been on a market for a period of time generates a trail of data. The basic premise of the data-driven innovation is that this trail of data is captured. Once the data is stored, it can be analyzed and the produced results could be used to foster innovation. The nature of the data collected for a family of eight products (Product 1 through Product 8) is illustrated in Table 1. For each product in the family, three types of features have been defined, product, customer, and market features.

	Product Feature					Customer Feature					Market Feature			
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
Product 1	p					c			c	c	m		m	m
Product 2				p	p		c	c					m	
Product 3	p		p			c			c		m	m		m
Product 4					p		c	c		c			m	
Product 5			p						c		m		m	m
Product 6		p			p	c		c		c		m		
Product 7	p						c					m		m
Product 8		p		p		c			c	c	m		m	

Table 1. Extended set of product features.

Of course, the product features can be considered at various levels, ranging from subassemblies (e.g., subassembly number, subassembly cost) through parts to component features (e.g., slot type, material type, and volume). Such data is usually collected for every product configuration sold.

Many companies had begun developing data warehouses to store information about current and potential customers (e.g., customer income, preferences, and purchase patterns) and market data (e.g., interest rates, weather, and trends). The width of the data (the type and the number of features) will likely grow in time. Such data will find a multitude of applications, including frequent product changes and innovation.

The Rise of Data-Centric Approach

There are at least three reasons behind the emergence of the data-driven innovation concept. First, the data-driven innovation stems from the information age when massive volumes of data are generated. The advancement of sensory and data-collecting technology, e.g., Radio Frequency Identification (RFID), has made gathering large amount of data possible. The rise of internet business has pushed seller-customer relationships to a virtual level in terms of handling transaction details (e.g., product specifications, shipping methods), customer preferences (e.g. the time spent on a particular webpage), and customer responses (e.g., feedback questionnaires). The time is ripe to move from data storage and data retrieval to data analysis and fostering data-inspired innovation. Second, intelligent algorithms and tools to process data and extract knowledge have been proposed and many of them have been proven to generate useful results. Data mining and evolutionary computation algorithms make the forefront of the list of candidates. Third, some companies with a longer innovation history, such as Apple Computer and 3M, have a lead-time and resource advantage over companies just entering the innovation trend. Hoffman *et al*. (1998) concluded that many formal and structured approaches developed for larger companies are not applicable for small firms. Data-driven innovation, however, does not narrowly constraint itself to be applicable to a particular industry or a specific size of the entity. Rather it is intended to be a common approach meeting the needs and targets of different companies.

Iterative Data Exploration

The fact that the conventional wisdom indicates some people are naturally more innovative than others and thus the ability to innovate can not be taught nor trained does not capture the entire picture of how innovation is formed. Indeed, some characteristics or personal traits, such as creativity and imagination, are beneficial in triggering innovation and it is assumed that innovators share at least some of these qualities. For example, innovators are open-minded to the possibilities of product opportunities and are able to balance them with customer needs and corporate strategy to create innovative products (Vogel *et al.* 2005). However, acknowledging the benefits of having these inherently tacit qualities to innovation does not provide much meaning as the diffusion of these qualities to other individuals is difficult. The ability to innovate, like many other skills, should be taught. If the data contains some meaningful implications to innovation, it is important to locate them. Moreover, different individuals can come up with different innovative ideas using the same data (see Fig. 1). The large innovation space depicted in Fig. 1 is dynamic, evolving, and it encompasses innovative ideas discovered by different individuals.

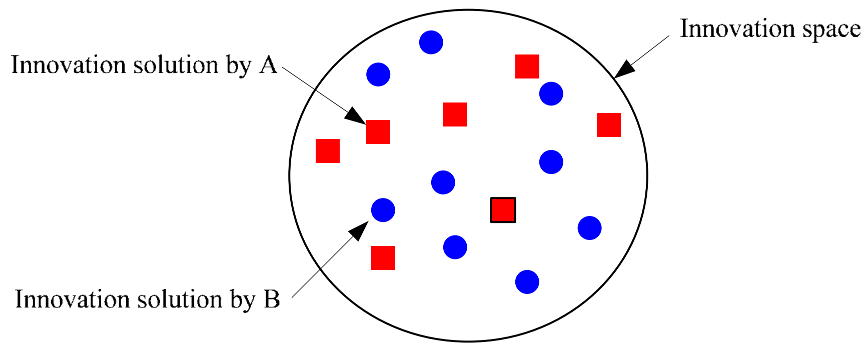

Figure 1. Innovation space.

The data-driven innovation by no means automatically produces innovation nor is it the only way to generate innovation. However, it is a pragmatic and systematic approach toward innovation. Assisted by a human it iteratively synthesizes heterogeneous data. The interpretation of information acquired by analyzing data is the key to successful innovation endeavor. It is important to note that the process of data-driven innovation is dynamic and it is scenario dependent. Direct inferences and conclusions made from other so-called "best practices" running the danger of observation of bias.

Companies adopting data-driven approach should be aware of two different initial assessments. If the problem to be solved is well defined or the requirements are well-documented, the decision flow presented in Fig. 2(a) is more appropriate. For the opposite scenario, the decision flow in Fig. 2(b) applies.

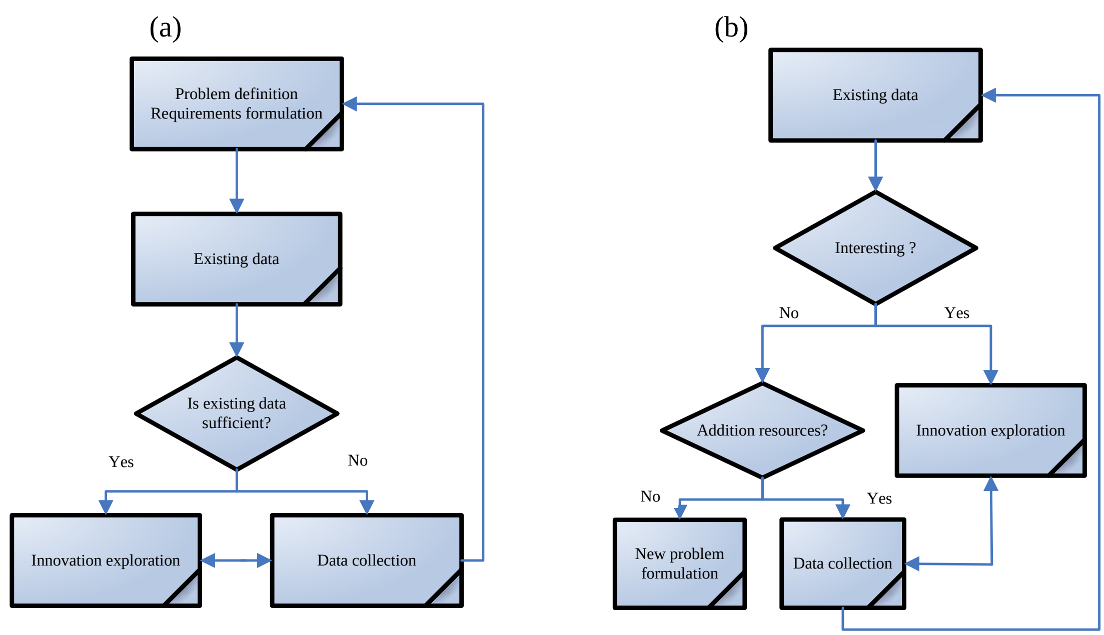

Figure 2. Two process flow in data and requirement matching.

Data analysis is critical in data-driven innovation. Traditional statistical methods have been frequently used to analyze data, however, stringent assumptions of some models limit their applicability. Data mining and evolutionary computation have created new possibilities for discovery of interesting patterns, trends, and associations from the data. A brief overview of the two disciplines is discussed next.

(1) Data mining

Data mining, an integral part in the process of knowledge discovery in database (KDD), is the process of generating useful information from raw data (Tan *et al*. 2005). Many successful applications in business, biosciences and engineering have been reported (Ganguly and Gupta 2004). It is an interdisciplinary field incorporating machine learning, artificial intelligence, and statistics. It is largely based on supervised learning and unsupervised learning. In supervised learning, a training data set with inputs and an output is used to build the learning function. In unsupervised learning the output class is not defined a priori. Classification and prediction fall under supervised learning, while clustering is widely used in unsupervised learning. Algorithms such as neural networks (NN), decision tree, support vector machine (SVM), and k-means clustering are widely applied. Readers interested in more details may refer to Tan *et al.* (2005) and Larose (2005).

(2) Evolutionary computation

Evolutionary computation covers the study of the foundations and applications of heuristics algorithms based on the principles of natural evolution. Examples of techniques include genetic algorithms and ant colony optimization (De Jong 2006). From the innovation perspective, the most promising innovative idea should be the "best-fit" for the market, similar to the concept of evolutionary computation looking for the "best-fit" to the environment. The link between innovation and evolutionary computation has not been built yet, however, there appear to be a natural match between the two.

3. Implications of Product Data

To illustrate how product's data can be utilized in innovation assume the following result derived from mining sales data with a decision-tree algorithm:

IF Product = "A" AND Sales Region = "East" AND Sale period = "Fourth Quarter" THEN Profit = "50%"

To determine the root cause of this outcome (Profit), the first three attributes are good initial candidates for further analysis. Although the actual cause is not known at this point, a few good starting points can be helpful in analyzing problems with large number of potential causes. Once the "golden-nugget" is found, capitalizing on it will produce benefits.

Consider another case with product data represented by the tree structures in Fig. 4. The trees refer to the same product but with different design modifications in subsequent periods. The root node at the top level represents the product and the nodes at the intermediate levels are the functional groups. The leave nodes at the very bottom are the parts.

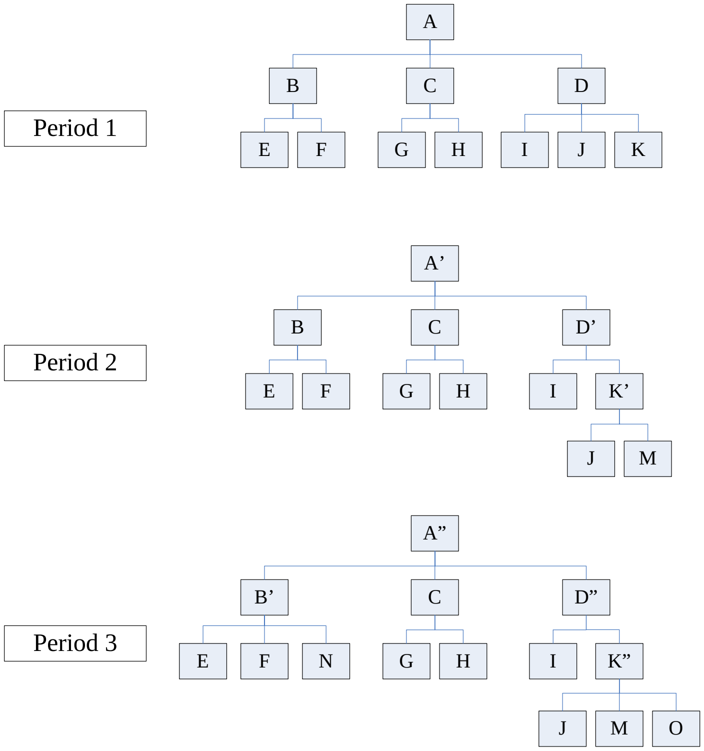

Figure 4. Evolution of product configurations.

Numerous implications can be derived from observing the three tree structures, for example:

(1) Functional group D (with its variants of D’ and D”) has undergone redesign at every period. Do these redesign efforts simply reflect the market trends or are they conducted to meet customer requirements in terms of functionality and quality?

(2) The design functional group C remained unchanged. Are the current designs satisfying customer needs? Are there improvements to make?

The answers to these questions are not trivial and additional data may be needed such as customer feedback, manufacturing data, and transaction data from sales department for more detail analysis. It is likely that additional data need to be collected to generate new ideas or new direction for improvements. The novel design of click-wheel of IPod by Apple is a good example. Before the initiation of this innovative design, most MP3 players used buttons for users to interact with the internal software. The customers then may have not pointed to any “problems” with this function because they did not have better reference or ideas to improve the design. This example shows how product data can be used to stimulate innovation initiatives in places that the customers value the most. It should be noted that the product data changes over time, so should innovation initiatives to reflect the current trend. Direct human interpretation of large industrial databases is time-consuming, if possible at all. Therefore, sophisticated and tailor-made intelligent methods should be applied instead.

Understanding how product data contributes to innovation is essential but how to use it poses another challenge. Data may be static but the process to find and spark innovative ideas is dynamic, so is the innovation itself. A system synthesizing heterogeneous data should be able to address these issues and provide fast response to ad-hoc user defined queries.

4. The Data-Inspired Innovation Model

There is a view in the literature purporting that innovation is customer driven. This view is usually illustrated with examples from pre-industrial era where a customer was getting a custom made product. It is being argued that the variety of customer specifications and possibly product makers has made some product innovative. It is argued that the emergence of industrial corporations has changed the individual customer requirements and customized product delivery aspect by the creation of marketing departments that have somehow transformed (e.g., averaged) the requirements. There is no doubt that the pre-industrial era scenario could be replicated in the information age where customer requirements are captured with great speed and accuracy. In fact, some industries (e.g., high end fashion industry) may be practicing this model. However, it is doubtful that this approach could be generalized to many industries.

What innovation-fostering approach would then work today? Recall that the pre-industrial era example involved two parties a customer and a product maker (e.g., a shoe maker). The innovation literature tends to emphasize innovation inspired by the customer. What about the product maker who was realizing the customer requirements? Some could argue that s/he has made valuable, if not greater than the customer, contributions to the innovation. The data-driven model of fostering innovation discussed next transforms the two-party example into a practical innovation model, called here data-inspired innovation (DII) model.

The data-inspired innovation (DII) model is applicable to the product redesign scenario. Redesign of any product or product family brings up the issue of innovation. As illustrated in Fig. 1 any product that has entered the market leave a data trail. This data can be analyzed and used for various purposes. For example, that data may support solving problems such as product modularity and product complexity reduction.
The data collected over the product life-cycle may be also mined for traits related to innovation. Some of those traits could make valuable requirements in support of innovation. Fig. 5 shows the requirements created by "product stakeholders" ranging from sales through manufacture to the innovation ideas provided by the customers and service.

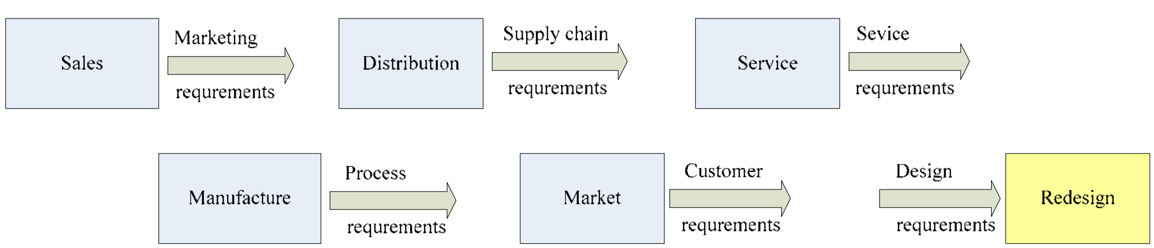

Figure 5. Innovation-inspiring requirements generated across the product life-cycle.

Some of the innovation requirements can be generated from the routinely collected data, while other may call for an extra effort, called data farming (Kusiak 2005). Data farming aims at defining parameters that are relevant to the problem at hand. In this case, the data farming methodology can be used to define and collect new data points across the product-life cycle that ought to be collected so that the products could be redesigned for innovation. In same cases measures leaser than the product redesign, e.g., a new product configuration, a new feature added, could suffice to make the product more innovative.

5. Conclusion

Many consider innovation as the driving force of modern economy. Corporations are ready to commit significant resources to enhance their ability to innovate. The research community has studied various aspects of innovation with some success. In this paper the concept of data-driven innovation was proposed. Traditional data analysis has focused on data manipulation rather than knowledge exploration from the data. Much valuable information is embedded in the data and potential benefits can be paramount if the information is extracted and utilized. Constraints, trends, and implications of the existing process are often in the data and any innovative idea or solution needs to be synchronized with these parameters.

The data-inspired innovation model presented in the paper emphasizes the importance of triggering innovative ideas from the data.

References

De Jong, K.A. (2006), *Evolutionary Computation: A Unified Approach*, MIT Press, Cambridge, MA.

Hoffman, K., Parejo, M., Bessant, J. and Perren, L. (1998), Small firms, R&D, technology and innovation in the UK: A literature review, *Technovation*, Vol. 18, No. 1, pp. 39-55.

Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., and Karthik, R. (2005), Shaped-based searching for product lifecycle applications, *Computer-Aided Design*, Vol. 37, No. 13, pp. 1435-1446.

Klepper, S. (1996), Entry, exit, growth, and innovation over the product life-cycle, *American Economic Review*, Vol. 86, No. 3, pp. 562-583.

Kusiak, A. (2005), Data Farming: A Primer, *International Journal of Operations Research*, Vol. 2, No. 2, pp. 48-57.

Kusiak, A. (2005), Innovation Science: A Primer, Working Paper ISL_11_2005, Intelligent Systems Laboratory, The University of Iowa, Iowa City, IA.

Ganguly, A. R. And Gupta, A. (2004), Data mining tehchnologies and decision supprot systems for business and scientific applications, *Encyclopeida of Data Warehousing and Mining*, Idea Group Publishing, Cambaridge, MA

Gopalakrishnan, S. and Damanpour F. (1997), A review of innovation research in economics, sociology and technology management, *Omega,* Vol. 25, No.1, pp. 15-28.

Parthasarthy, R. and Hammond, J. (2002), Product innovation input and outcome: moderating effects of the innovation process, *Journal of Engineering and Technology Management*, Vol. 19, No. 1, pp. 75-91.

Piana, V. (2003), Innovation, http://www.economicswebinstitute.org.

Tan, P-N, Steinbach, M. and Kumar, V. (2005), *Introduciton to Data Mining*, Pearson Eduction, Boston, MA

Tang, C.-Y. and Kusiak, A. (2006), The Frontiers of Innovation: Review and Extensions, Working Paper ISL_03_2006, Intelligent Systems Laboratory, The University of Iowa, Iowa City, IA.

Witten, I.H. and Frank, E. (2005), *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, San Francisco, CA.

Troyer, L. (2005), Personal communication, Department of Sociology, The University of Iowa, Iowa City, IA.

Vogel, C.M, Cagan, J., and Boatwright, P. (2005), *The Design of Things to Come*, Wharton School Publishing, Upper Saddle River, N.J.