Data Mining and Innovation Science

Andrew Kusiak
Mechanical and Industrial Engineering, 3131 SC
The University of Iowa
Iowa City, IA 52242-1527
andrew-kusiak@uiowa.edu

Abstract

A newly introduced product or service becomes an innovation after it has been proven in market. No one likes the fact that market failures of products and services are much more common than commercial successes. The ideas introduced in this paper are applicable to the evaluation of the innovativeness of planned introductions of design changes, products, and services. In fact, blends of products and services could be the most promising way of bringing innovations to the market. The most important toll gates of innovation are a generation of new ideas and their evaluation. People have limited ability to generate and evaluate a large number of potential innovation alternatives. The proposed approach provides a number of such alternatives and evaluates them from the market perspective.

1. Introduction

The study of innovation – the development of new knowledge and artifacts – is of interest to engineering, business, social and behavioral sciences, and spans sociology, history, philosophy, economics, psychology, and political science. Innovations transform economies (e.g., the knowledge-based Silicon Valley economy was created from California's agricultural economy). Innovations alter global relations (e.g., the creation of international treaties due to developments in nuclear technologies) and produce new structures of social control (e.g., the creation of international regulatory agencies to oversee pharmaceutical industries and new state bureaus charged with investigating cyber-crime). Innovations change the day-to-day lives of individuals (e.g., the introduction of new biopharmaceutical discoveries affect life expectancy and the quality of life).

Though many innovation studies have been published, the literature on innovation is filled with myths and inconclusive research findings. Innovation is often discussed based on experiences specific to a particular case study. For example, innovation undertakings at companies such as Procter and Gamble and Apple Computer have been broadly studied. However, is not known to what degree these findings would produce similar results in other corporations. The need to create innovation science is apparent as outlined in [1].

What Is Innovation?

Innovation is an iterative process aimed at the creation of new products, processes, knowledge or services by use of new or even existing knowledge. Some prefer the terms "technology-based innovation" or "technological innovation" to emphasize the role of technology.

What Drives Innovation?

Companies use various means to reach out to customers to incorporate their needs into the product development process. Many researchers have suggested that companies use an incorrect approach and incorrect measurements when consulting with customers. Ulwick [2] pointed out that companies should not expect solutions to be offered by potential customers; rather, they should ask them about the desired product's characteristics. He argued that customers may only know what they have experienced and may have a limited frame of reference when suggesting innovative ideas. In addition, companies that link their products too closely to their customers may end up creating incremental innovation. Veryzer [3] emphasized the need for caution with customer input and pointed out the importance of discontinuous product development, e.g., the customer's input should be introduced later in the project. Christensen [4] stated that customers may overemphasize the product's functionality.

2. Innovation Process

Nambisan and Sawhney [5] discussed three types of innovation intermediaries, each operating in its own landscape:

- Invention capitalist (iC),
- Innovation capitalist (IC), and
- Venture capitalist (VC).

Each of the three innovation landscapes follows the generic process model shown in Fig. 1. This model generalizes the steps used the invention capitalist approach outlined in [5].

Figure 1. Innovation landscape process model.

Each innovation intermediary performs the following five activities: search, evaluate, develop, refine, and connect, however, in a different risk scenario and cost landscape. This landscape determines the input to the search activity and the output of the connect activity. The input to the search activity for an invention capitalist includes predominantly inventions and ideas, and the goal is to connect companies with the inventions and ideas that are promising but not market-ready yet. For an innovation capitalist the inputs are market-ready ideas, and the goal is to connect companies with the market-ready ideas. A venture capitalist follows the same process (Fig. 1), where the input constitutes market-ready products, and the goal is to connect companies with the market-ready ventures.

Each of the three innovation landscapes involves different risks and costs. The cost-risk relationship between these landscapes is shown in the grid in Fig. 2.

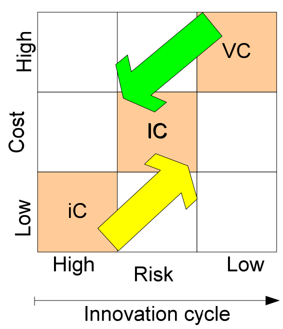

Figure 2. Cost-risk innovation grid.

The innovation capitalist (IC) optimizes the tradeoff between the cost of bringing market-ready ideas to market and the associated risk.

The transition from the invention square (iC) to the venture square (VC) in its simplest form is along the diagonal of the grid.

The arrow below the diagonal in Fig. 2 indicates the natural progression from the iC to the VC landscape. While the focus of business activities in recent decades has been on the VC quadrant, the arrow above the diagonal symbolizes the direction of focus needed to energize innovation for companies focused on venture capital driven innovation.

A company interested in innovation from outside sources needs to carefully balance the three different innovation landscapes. According to [5], basing innovation on the lower left square in Fig. 2 appears to be attractive to consumer products and markets populated with many different and relatively simple products. The top left right area may apply to companies that are science and technology driven, e.g., 3M and DuPont. Development cost of products manufactured by these companies is high, and therefore the innovation is likely to come from collaborating companies with significant human and capital resources. The innovation diffusion can be accelerated by management strategies moving companies from the upper right and lower left squares towards the center of the grid in Fig. 2.

3. Requirements-Guided Innovation

The customer perspective has been behind the creation and processes in the last two decades. This market focus has generally been reflected in the product's functions and form. Other commonly used attributes to attract customers and at the same time improve business performance have been quality, reliability, and cost.

The level of innovation I can be expressed as a function of requirements X, $I = F(X)$, involving various classes of requirements (X): function, form, surprise, culture, emotion, and experience (see Fig. 3).

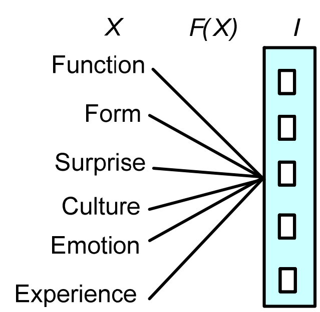

Figure 3: A tree of expanded requirements.

The list of requirements impacting innovation expands beyond function and form. In fact, the AND/OR tree representation allowing the inclusion of alternatives may be used to elicit and represent requirements [6]. The approach advocated in this paper calls for broadening the range of requirements over the traditionally considered ones (mostly function and form).

Understanding the breadth, content, and structure of the customer requirements is key to innovation. A customer of today purchases a product that meets her/his functional requirements (product personalization), but also seriously considers additional attributes such as surprise (e.g., unexpected product function), pleasure (e.g., driving a car), emotion, and so on.

Ultimately the increased level of innovation *I* has to translate into business benefits, e.g., increased market share.

4. Sources of Requirements

In the past two decades, the design of products and services has been largely driven by customers. After all, the customer buys a product or uses a service. The "customer-as-the-king" model was preceded by the "engineer-as-the-king" (often designer) model, in which technical experts made the decisions for the customer. The customer was expected to accept the offered product or service.

Both models of eliciting requirements have focused on the product and service functions. Product innovation calls for additional requirements, making it worthy of the label "innovative product." The sources of innovation-fostering requirements are much wider and they include:

- Customers. The information from the customers should be collected over the product's life-cycle rather than during a limited time frame. Processing that information and blending it with other sources of data and information could be the ultimate key to the success of the designed product.
- Domain experts. Though the importance of the voice of the engineer in forming requirements has been marginalized in the last few decades, it needs to be brought back and expanded when innovating. It is true that a customer is the one who ultimately pays for the product; however, he may not be aware of the possibilities that a new technology or a product/process combination may offer. A technologist may generate innovative features of a product.
- Legacy materials. All kinds of standard and digital libraries could be searched in the quest for innovation. The search would involve hypotheses, theories, innovation rules, and information about inventors and innovators. Data-mining algorithms could create previously unseen value in fusing data and information from various sources.
- Product life-cycle data. A product leaves a data trail over its life cycle. This is in addition to the information provided by the customers or experts before and after the product has entered the market. The volume of data collected can be large, e.g., imagine a database of cockpit and maintenance data collected over the useful life of an airplane. The product's lifetime data can deliver valuable knowledge leading to requirements spurring innovation.

Having outlined the role of requirements in innovation as a "data generator," the role of data mining in this exciting undertaking is obvious. It will be used to discover patterns leading to market acceptance of candidate solutions.

Without going into detail, many agree market relevance and market acceptance distinguish innovation from invention and creation (see Fig. 4).

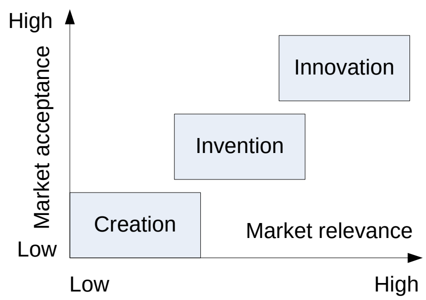

Figure 4. Relationship between creation, invention, and innovation.

The market determines whether a creation or an invention becomes an innovation. The market acceptance and relevance can be expressed in economic terms (e.g., market share, profit) or other metrics (e.g., social acceptance). Data mining is likely to play a key role in focusing on and pursuing creations and inventions that have a high likelihood of becoming innovations.

5. Innovation Evaluation

The interest in innovation is not new; however, it has become of particular interest in recent years due to numerous factors, including the increasing dynamics of a global economy. Next, examples of methods and tools for the evaluation of innovations are discussed.

5.1 Trial and Error Approach

A widely used approach to innovation is trial and error. Designers observe the consequences of the design choices made and learn from them. The advantage of the trial-and-error approach is that it is easy and everyone can use it. The major limitation of this approach is the lack of predictability of the outcome.

4.2 Lead User Study

The lead-user market research method is based on the concept that the need for new products, processes, or services is best understood by a few well informed uses, called lead users. This concept was introduced by von Hippel [7]. The lead users can be incorporated into a development process of a joint or new product, process, or a service with the company's developers. Herstatt and von Hippel [8] demonstrated in a case study that the lead-user approach was almost twice as fast as traditional ways of identifying promising new product concepts and less costly.

The lead-user method involves four major steps [7, 8]:

Step 1: Specifying product/service characteristics of interest to future customers.

Lead users of a product, process, or a service are persons who display two characteristics:

(*A*) They anticipate important marketplace trend(s),

(*B*) They have a good sense of the benefits offered by the purported solution.

Step 2: Identifying lead users.

Step 3: Engaging lead users in the development of product/service concepts.

Step 4: Testing the concepts developed by lead users in a sample market of typical users.

4.3 Innovation Networks

Innovation enables organizations to effectively compete [4] by supporting the innovation process, e.g., the idea generation phase, conversion, or diffusion phase (see Fig. 3). The need to innovate has resulted in renewed interest among research and corporate communities. Though numerous innovation studies have been published, myths and inconclusive research findings are quite common. Innovation is often discussed based on experiences specific to a particular case study. For example, innovation undertakings at companies such as 3M and Apple Computers have been broadly studied. However, is not known to what degree these findings would produce similar results in other corporations.

The most difficult issue is that of predicting the success of a product/service at an early stage of its development. The published literature does not provide any evidence that such a tool exists.

In recent years there has been an increased interest in innovation in networked environments, especially in the European literature. This could be due to the networked research environment promoted by the projects sponsored by the European Commission. In fact, the focus of some of these projects has been on studying collaboration, e.g., the ECOLEAD initiative (www.ecolead.org) involving over 20 partners from 12 countries and funded at the level of Euro 18M. Another measure of the growing interest in networked organizations is the recently established Society of Collaborative Networks, SOCOLNET (http://www.socolnet.org).

The emergence of domestic networks seeking customer-based information needs to be noted, e.g., http://www.ninesigma.net and erewards@e-rewards.net. Though the scope, functionality, and research value added by the commercial networks may be limited, the trajectory of using market information in the development of products/services is clear.

Chiffoleau [9] presented the results of a longitudinal ethnographic case study. A small cooperative implemented environmental-friendly viticulture in Southern France. The study stressed the involvement of domain experts beyond "traditional" leadership and management of "practice networks" by integrating these networks and linking diverse strategic positions to handle innovation challenges.

The synthesis approach to innovation in service and manufacturing was studied by de Vries [10]. The theory of Gallouj and Weinstein [11] was modified in order to consider the innovation trends in networked organizations and in the distributed services. The modification studied was based on several case studies.

Corporations attempt to improve their performance by engaging in radical or incremental innovation through partnerships and networking with other corporations. The simulation experiments reported by Gilberta *et al.* [12] showed the impact of various learning activities on innovation.

An issue of concern, especially for novices of collaborative networks, is that of handling intellectual property. Many will agree that research is needed to develop different models of handling a company's confidential information. A natural way of limiting the release of proprietary information is by using an open communication channels customer feature rather than technical product/service features.

Benkler [13] and von Hippel [14] used different terms to describe the involvement of the market in the innovation process, and both have stressed that handling intellectual property needs to be investigated in the future. In fact ways of handing issues related to intellectual property in a networked environment could indicate a measure of success. Some results of handling intellectual property issues have begun to emerge. For example, Henkel [15] discussed the results of a quantitative study (N = 268) of patterns of freely revealing firm-developed innovations embedded in Linux, an open source software. The author observed that corporations contributed (without obligation) their own developments to the Linux code. In return they elicited and received informal development support from other corporations. Though this open exchange of information would be unthinkable for traditionally minded managers, a part of corporate product development was performed in an open environment. The issue of intellectual property was addressed by selectively revealing information. A corporation would reveal, on average, about half of the code it had developed, while protecting the other half by various means. Revealing was strongly heterogeneous among firms. Analysis of reasons for revealing and of the type of revealed codes showed that the rationale for openness varied across corporations. The conflict between benefits and drawbacks of openness appeared to be manageable.

5. Proposed Innovation Framework

5.1 Solution Architecture

The key issue in innovation is an early evaluation of many possible solution alternatives. A traditional approach limits the number of alternatives due to the time and cost necessary to create and evaluate them.

The basic steps of the proposed three-phase approach are illustrated in Fig. 6 through Fig. 8 (see [16]). Fig. 6 illustrates Phase 1, where a training data set is generated. A design team develops a prototype model (or a few prototype variants) involving innovation features (called here an aspect). The initial prototype set (Fig. 6) is expanded by an evolutionary computation algorithm into a large prototype set evaluated in the test market producing a training data set Si, i = 1,.., n.

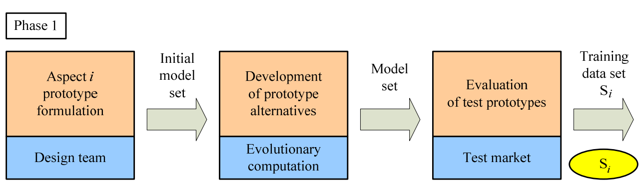

Figure 6. Aspect *i* learning mode.

In Phase 2, the training data sets S*i* for *n* aspects are integrated into a single data set S used to build a classifier shown in Fig. 7.

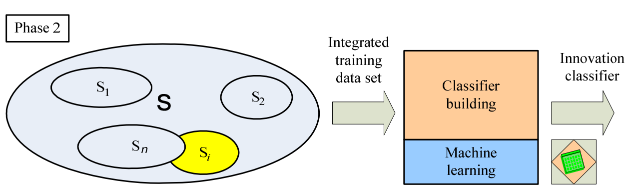

Figure 7. Learning from an integrated data set of *n* aspects.

In Phase 3, the classifier or an ensemble of classifiers built in Phase 2 is evaluated for accuracy and used to predict the success (e.g., innovation score) of the test configurations to be considered for further development (see Fig. 8).

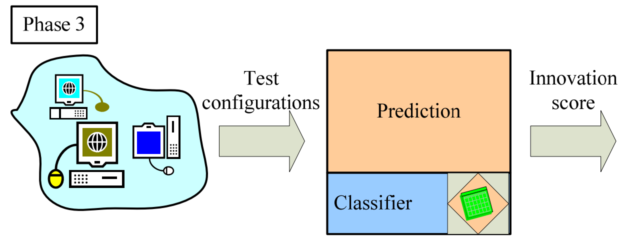

Figure 8. Prediction of innovation scores.

The three-phase approach presented involves feedback loops not shown in Figures 6 through 8. For example, the configurations evaluated at the end of Phase 3 could be introduced to the training set used as input in Phase 2.

Evolutionary computation, in particular a genetic programming algorithm (GP), appears to naturally match the methodology gap of Phase 1. Generation of solution (configuration) alternatives, though realized by the genetic programming algorithm, requires an innovation evaluation (fitness) function. In the proposed research, a data-mining scheme is presented to develop a classifier for the evaluation of a large number of the expert and GP-generated configurations (solutions). The classifier will be extracted from a training data set produced from the intermediate solution set. The intermediate solution set will usually be larger that the initial solution set, however, much smaller than the expanded solution set.

5.2 Novelty of the Proposed Framework

The novelty of the research presented is realized at two different levels. The first level is the most challenging part of the proposed research, the design of a system for enabling innovation. The sources of data, as well as interoperability among all constructs and algorithms, will be established. The proposed solution to test innovations will be known as the Living Laboratory of Innovation Discovery (LIVLID) outlined in [17]. The framework is a step towards the realization of innovation science [1, 18].

6. Conclusion

A framework was outlined for innovation based on the requirements elicited from multiple sources. Like innovations generated through market success, any development process has to target the market-expected requirements. With the abundance of data in the cyber world, new ways to analyze and use the data are needed. The collected data and requirements are refined and analyzed by tools and human resources all assembled as the Living Innovation Laboratory in service of innovation.

7. References

[1] A. Kusiak, Innovation science: A primer, *International Journal of Computer Applications in Technology*, 2007, Vol. 28, No. 2-3, pp. 140-149.

[2] A.W. Ulwick, Turn customer input into innovation, *Harvard Business Review*, 2002, Vol. 80, No. 1, pp. 91-97.

[3] R. Veryzer, The roles of marketing and industrial design in discontinuous new product development, *Journal of Product Innovation Management*, 2005, Vol. 22, No. 1, pp, 22-41.

[4] C. Christensen, The innovator's dilemma: When new technologies cause great firms to fail, *Harvard Business School Press*, Boston, MA, 2007.

[5] S. Nambisan and M. Sawhney, A buyer's guide to the innovation bazaar, *Harvard Business Review*, 2007, Vol. 85, No. 6, pp. 109-118.

[6] A. Kusiak and E. Szczerbicki, A formal approach to specifications in conceptual Design, *ASME Transactions: Journal of Mechanical Design*, 1992, Vol. 114, No. 4, 659- 666.

[7] E. Von Hippel, Lead users: A source of novel product concepts, *Management Science*, 1986, Vol. 32, No. 7, pp. 791-805.

[8] C. Herstatt and E. von Hippel, From experience: Developing new product concepts via the lead user method: A case study in a "low tech field", *Journal of Product Innovation Management*, 1992, Vol. 9, pp. 213-221.

[9] Y. Chiffoleau, Learning about innovation through networks: The development of environment-friendly viticulture, *Technovation*, 2005, Vol. 25, No. 10, pp. 1193-1204.

[10] E.J. De Vries, Innovation in services in networks of organizations and in the distribution of services, *Research Policy*, 2006, Vol. 35, No.7, pp. 1037-1051.

[11] F. Gallouj and O. Weinstein, Innovation in services, *Research Policy*, 1997, Vol. 26, pp. 537–556.

[12] N. Gilberta, P. Ahrweilerb, and A. Pykac, Learning in innovation networks: Some simulation experiments, *Physica A: Statistical Mechanics and its Applications*, 2007, Vol. 378, No. 1, pp. 100-109.

[13] Benkler, Y., *The Wealth of Networks*, Yale University Press, New Heaven, CT, 2006.

[14] Von Hippel, E., *Democratizing Innovation*, MIT Press, Cambridge, MA, 2005.

[15] J. Henkel, 2006, Selective revealing in open innovation processes: The case of embedded Linux, *Research Policy*, 2006, Vol. 35, No. 7, pp. 953-969.

[16] A. Kusiak, Innovation of products and services: Bridging world's economies, *Proceedings of the 19th International Conference on Production Research*, ICPR 19, Valparaiso, Chile, August 2007, pp. 1-6.

[17] A. Kusiak, Innovation: The living laboratory perspective, *Computer-Aided Design and Applications*, 2007, Vol. 4, No. 6, pp. 863-876.

[18] A. Kusiak and F.A. Salustri, Computational Intelligence in Product Design Engineering: Review and Trends, *IEEE Transactions on Systems, Man, and Cybernetics*: Part C, 2007, Vo. 37, No. 5, pp. 766-778.