International Journal of Information Technology & Decision Making
Vol. 13 (2014)

DOI: 10.1142/S0219622014500266

AN EXTENSION COLLABORATIVE INNOVATION MODEL IN THE CONTEXT OF BIG DATA

XINGSEN LI[*,¶], YINGJIE TIAN[†,‖], FLORENTIN SMARANDACHE[‡,**]
and RAJAN ALEX[§,††]

[*]*School of Management, Ningbo Institute of Technology*
Zhejiang University, Ningbo 315100, P. R. China

[†]*Research Center on Fictitious Economy and Data Science*
CAS, Beijing 100190, P. R. China

[‡]*Mathematics and Science Department, University of New Mexico*
Gallup, NM 87301, USA

[§]*School of Engineering and Computer Science*
West Texas A&M University, Conyon, TX 79015, USA
[¶]*lixs@nit.zju.edu.cn*
[‖]*tianyingjie1213@163.com*
[**]*smarand@unm.edu*
[††]*ralex@wtamu.edu*

Published 2 December 2013

The processes of generating innovative solutions mostly rely on skilled experts who are usually unavailable and their outcomes have uncertainty. Computer science and information technology are changing the innovation environment and accumulating Big Data from which a lot of knowledge is to be discovered. However, it is a rather nebulous area and there still remain several challenging problems to integrate the multi-information and rough knowledge effectively to support the process of innovation. Based on the new cross discipline *Extenics*, the authors have presented a collaborative innovation model in the context of Big Data. The model has two mutual paths, one to transform collected data into an information tree in a uniform basic-element format and another to discover knowledge by data mining, save the rules in a knowledge base, and then explore the innovation paths and solutions by a formularized model based on *Extenics*. Finally, all possible solutions are scored and selected by 3D-dependent function. The model which integrates different departments to put forward the innovation solutions is proved valuable for a user of the Big Data by a practical innovation case in management.

Keywords: Extension innovation model; Big Data; data mining; Extenics; knowledge management.

1. Introduction

In the past few decades, a large number of scholarly efforts and theories on collaborative innovation have been developed, and many approaches have contributed to reveal the nature of innovation process in various degrees.[1,2] Nevertheless, the innovation process still largely remains a black box.[3] Although each theory can explain some of the mechanisms behind innovations, the general mysteries behind innovation process are still far from resolved. Most of the innovation models make use of a group of experts and depend on personal intelligence which would be subject to limitations of individual opinions, and therefore they keep out of step from the rapidly changing information and knowledge era.

We live in an era where a remarkable amount of new information and data is accumulated everywhere. Because of the massive amount of data that is generated almost everywhere, new tools need to be developed in order to manage and analyze the data, especially in the field of management.[4] The information environment is rapidly and continuously changing and uncertain due to global competition, information explosion, and advances in new technologies. The web and other information technologies (ITs) combined with business and lives have accumulated huge data and information. Information and other technologies are successfully bypassing the main obstacle to technological advance, and when technology support net is fully established and fixed, innovation is not a free and autonomous process of applied creativity but is technically, economically and politically subservient to the "holders and owners" of the support net.[5]

Big Data[6,7] is a collection of data sets. It is so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.

Data is accumulating from almost all aspects of our everyday lives that it becomes huge and multi-structured and has hidden useful information. The era of Big Data is real and is going to stay with us.

The challenges with Big Data include capture, curation, storage, search, sharing, transfer, analysis, and visualization (http://en.wikipedia.org/wiki/Big_data). Big Data provides materials for mining hidden patterns to support innovation[8] mostly by data mining.[9,10] The interaction research with Big Data support methods for innovation is rare at present.[11] Knowledge discovered by data mining is novel and quantitative.[12,13] However, it still lacks a uniform knowledge management model to support the innovation process effectively.[14]

Extenics (formerly referred as Extension Theory) focuses to solve contradictory problems by formalization methods based on the concepts of matter-element and extension set.[15,16] Extenics uses a uniform three-dimensional (3D) matrix to express information and knowledge and utilizes extension transformation to represent properties of things and indicates things with certain attribute that can be changed into things without such attribute. It provides a new view for understanding the process of innovation. But it needs more support of IT, especially Big Data.

To improve the quality of collaborative innovation by objective hidden knowledge from Big Data, we propose a new innovation model combining IT and Extenics. The rest of this paper is organized as follows. Section 2 discusses the existing innovation methods and problems. Section 3 provides a systemic model for preparing the input for innovation from Big Data. Section 4 presents a framework and process about utilization of Extenics to generate innovative ideas or solutions with a new method to score the solutions. Section 5 gives an application of the proposed Extenics to a real-world problem and Sec. 6 suggests future research direction for our work.

2. Literature Review

There are a number of factors that affect the quality of collaborative innovation including internal factors such as the will to change, the attitude to new things, the thinking model to overcome habitual domains,[17] and external factors, such as the information, data, team work, and the policy. Among them creative thinking model[18] and IT are the most important.[19]

Creative thinking mostly relies on individuals. It cannot be understood using a single simple model and it involves multiple complex processing operations. The operation of multiple processes, multiple strategies, and multiple knowledge structures makes it difficult to understand creative thinking process.[20] However, effective creativity execution depends on the knowledge available and the strategies people employ in executing these processes.[21]

The operation of multiple processes, multiple strategies, and multiple knowledge structures makes it difficult to formulate an understanding of innovation.[20] Declarative knowledge, factual, information, and cognitive schema are commonly held to be involved in most forms of creative thinking. Information and communication technology tools are likely to provide new innovation approaches and effective means to support such new innovation processes, such as classification algorithm selection in multiple criteria decision making.[22] The new approaches for innovation will find their wide application in industry.[23]

A lot of innovation methods make use of various approaches to stimulate innovation, such as individuals or group of experts/team members, along with numerous brain storming sessions involving both financial and human resources, and has severe dependence on personal intelligence; it would be subject to limitations of individuals themselves,[24] thereby keeping out of step from the rapidly changing information and knowledge environments.

To support innovation process, there are some tools that one can be deployed. There are eight core processes for the effective execution of innovation.[21] They are: (a) problem definition, (b) information gathering, (c) information organization, (d) conceptual combination, (e) idea generation, (f) idea evaluation, (g) implementation planning, and (h) solution monitoring. Effective execution of these processes, in turn, depends on people applying requisite strategies during process execution and having available requisite knowledge and it needs to be refined in detail for practical use.

TRIZ was developed to resolve contradictions in technological inventions, with a set of 40 inventive principles and later a matrix of contradictions which indicates 39 system factors.[25] It is useful in several specific fields, such as mechanics and electronics, but limited in other fields.

One main problem presented in many existing approaches is that they have applied a deductive approach by attempting to abstract common features from historical instances to obtain general rules for invention. Although this "expert system"-like approach is not without its merit, in reality, it is not realistic, because there are so many conditions and variables to match, and for each condition or variable, there are so many possible values to compare, so it may be computationally infeasible. On the other hand, although generative approach has been proposed by some authors, there has been a lack of effective ways of generating all possible innovative solutions.

Extenics focus on solving incompatible problems by formularized methods both in management and engineering. Zhou and Li[26] put forward an Extenics-based enterprise-independent innovation model and its implementation platform. Declarative knowledge, factual information, and cognitive schema are commonly involved in most forms of complex performance including innovation.[27] By integrating methodology knowledge and information, we need a theory to guide the generation of innovative solutions from Big Data.

Big Data is the next frontier for innovation, competition, and productivity[8]; it can help to better capture, understand, and meet customer needs.[28] Kou and Lou[29] proposed a hierarchical clustering method that combines multiple factors to identify clusters of web pages that can satisfy users' information needs. It is a very important source to knowledge[30,31] and other new discoveries.[32] But the question is: how to use Big Data to support collaborative innovation effectively? The data preparation process for innovation still remains a black box. Data is big enough but the methods to handle information and knowledge are very limited. Moreover, models listed above pay little attention to data analysis during innovation processes; the methods to score innovation solutions are mostly qualitative and we need more quantitative methods.

The Big Data and information is so huge that they are beyond human mind's processing capability. So it is time to implement collaborative innovation effectively in Big Data era. It is necessary to explore the high-efficiency models that would fill up the gaps with innovation process and new methodologies. We attempt to use Extenics to bridge the innovation process with data technology and management in this work.

3. Big Data Preparation for Collaborative Innovation

3.1. *Data collection based on Extenics*

Innovation process needs data and knowledge, both explicit and tacit. There are two main sources for collecting data: internal source, such as management information system, local database, tables or other forms of flat files, and external source, such as the internet, public databases, and data from other companies with similar goals.

There is huge quantity of data and information growing dramatically. How to choose the proper data set and process is a challenging problem.

Basic-element theory describe the matter (physical existence), event, and relations as the basic elements for all the information — "matter-element," "event-element," and "relation-element." Basic element is an ordered triad composed of the element name, the characteristics, and its measures, denoted by $R = (N, c, v)$ as matter-element, $I = (d, b, u)$ as event-element, and relation-element as $Q = (s, a, w)$.[16] As the matter-element $R = (N, c, v)$ is an ordered triad composed of matter, from its characteristics and measures, we can develop new concepts as the extensibilities of one of the three sub-elements in the triad.

Multiple characteristics are accompanied with multidimensional parametric matter-element and can be expressed as:

$$M(t) = \begin{bmatrix} O_m(t), & c_{m1}, & v_{m1}(t) \\ & c_{m2}, & v_{m2}(t) \\ & \vdots & \vdots \\ & c_{mn}, & v_{mn}(t) \end{bmatrix} = (O_m(t), C_m, V_m(t)).$$

A given matter has corresponding measure about any characteristic, which is unique at nonsimultaneous moments.

Further, characteristics of matters can be divided into materiality, systematicness, dynamism, and antagonism, which are generally called matters' conjugation. According to matters' conjugation, a matter consists of the imaginary and real, the soft and hard, the latent and apparent, and negative and positive parts,[16] which are explained below:

(1) *Non-physical part and physical part*

In terms of physical attribute of matter, all matters are composed of a physical part and a non-physical part. The former is referred to the real part of matter and the latter is referred to the *non-physical* or virtual part of matter. For example, a product's entity is its real part, while its brand and reputation are its *non-physical* parts. The empty space is a cup's *non-physical* part, while the ceramic cup itself is its *physical* part.

(2) *Soft part and hard part*

Considering a matter's structure in terms of the matter's systematic attribute, we define the matter's components as the hard part of matter, the relations between the matters and its components as the soft part of the matter. In the old saying, "Three heads are better than one," the three persons are the hard parts and the cooperation relationship is the soft part. A good soft part leads to a good result.

The matter's soft part has three types of relation: (1) relations between the matter's components; (2) relations between the matter and its subordinate matters, and (3) relations between the matter and other matters.

(3) *Latent part and apparent part*

Considering matter's dynamic property, we trust that any matter is changing. A disease has its latent period, a seed has its germination period, and an egg can hatch into chicken at a certain temperature after a certain time, and so on. The matter's latent parts and apparent parts exist synchronously.

The latent part of some matters may become apparent under certain conditions; for example, a student now in class may become a teacher after 10 years. There must be a criticality in the process of reciprocal transformation between latent parts and apparent parts.

(4) *Negative part and positive part*

In terms of antithetic properties of matters, all matters have two-part antithetic properties. The part producing the positive value to certain characteristic is defined as the positive part, and the part producing the negative value is defined as the negative part.

For example, in terms of profit, an employees' welfare department, a kindergarten, publicity departments, etc., have negative measure of profits, being the negative parts of the company, but these parts will improve employees' job enthusiasm and promote company's reputation, so they are the "advantageous" parts of the company.

Conjugate analysis and basic-element theory is a guide for us to collect data and information in a systematic way. Denote physical part as ph, non-physical part as nph, soft part as s, hard part as h, apparent part as a, latent part as l, positive part as p, negative part as n, matter-element as M, event-element as E, and relation-element as R. Using the notations, for example, *the physical part of relation* can be denoted as R_{ph}. Accordingly, we form a detailed data collecting list as showing in Table 1.[19]

Innovation activities have their goals and conditions. The purpose of data collection and processing is to solve problems about how to get from the conditions to the goals. Therefore, the data we collected should also be relative to the goals.

It can be seen from the above definition that there are three paths for the process of transformation between the element (such as "positive" and "negative"), field and criteria. Denote field as F, criteria as C, and element as EL, similarly, we denote goal as g, conditions as c, and pathways as pa. Accordingly, the data and information we collect will include such three paths.

Table 1. Data collection list based on basic-element theory and its conjugate analysis.

Basic-element type	Materiality		Systematicness		Dynamic		Antithetical	
	Physical part	Non-physical part	Soft part	Hard part	Apparent part	Latent part	Positive part	Negative part
Matter-Element	M_{ph}	M_{nph}	M_s	M_h	M_a	M_l	M_p	M_n
Event-Element	E_{ph}	E_{nph}	E_s	E_h	E_a	E_l	E_p	E_n
Relation-Element	R_{ph}	R_{nph}	R_s	R_h	R_a	R_l	R_p	R_n

Table 2. Data collection list from the view of the goal on certain business.

	Goals	Conditions	Pathways
Field	F_g	F_c	F_{pa}
Criteria	C_g	C_c	C_{pa}
Element	EL_g	EL_c	EL_{pa}

Goals and conditions can be matter, event, or relations between matters and events which can be represented with basic elements. So each cell in the Table 1 can be a basic element in next level of the information tree.

$$\begin{aligned}\text{Big Data} &= F + C + \text{EL},\\ \text{EL} &= \text{EL}_g + \text{EL}_c + \text{El}_{\text{pa}} = M + E + R,\\ M &= M_{\text{ph}} + M_{\text{nph}} + M_s + M_h + M_a + M_l + M_p + M_n.\end{aligned}$$

All the contents in Table 2 consist of Big Data of certain business. From Tables 1 and 2, we can get a systematic cube for collecting data and information for innovation.

3.2. *Data processing paths*

The Big Data preprocess chart is shown in Fig. 1.

There are two main paths located in four levels. One path is to extract data from the database, or use web crawler to collect information from the web, then transform and filter it into data mart, and finally use data mining to discover primary knowledge. Another way is to collect documents and build an information cube by human–computer interaction, then save it as basic-elements. Using extension

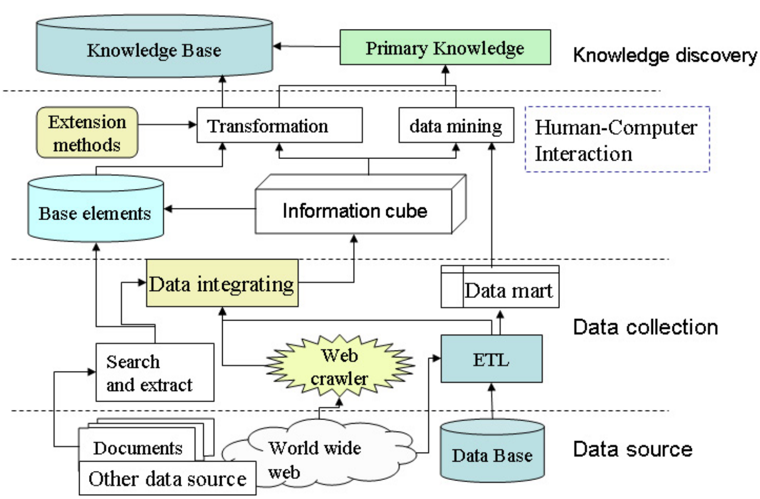


Fig. 1. Big Data collection and processing.

transformation methods, we transform the basic-elements into a knowledge base. Basic-elements and knowledge base will be the input of collaborative innovation model.

3.3. *Data transformation methods*

There are five basic transformation methods in Extenics, which can be used for information transformation by the change of a matter's object, attribute, or value.

(a) Substitution transformation
As to basic-element $B_0(t) = (O(t), c, v(t))$, if there is certain transformation T that transforms $B_0(t)$ to $B(t) = (O(t), c, v(t))$, i.e., $TB_0(t) = B(t)$, then the transformation T is referred to as substitution transformation of the basic-element $B_0(t)$.

(b) Increasing/decreasing transformations
An increasing transformation refers to the increase of certain attributes of the element. For example, as with matter-elements $M_0 = (\text{table A}_1, \text{height}, 0.8\,\text{m})$, $M = (\text{chair A}_2, \text{height}, 0.5\,\text{m})$, M is an increasable matter-element of M_0, we make $TM_0 = M_0 \oplus M = (\text{table A}_1 \oplus \text{chair A}_2, \text{height}, 1.3\,\text{m})$, then T is an increasing transformation of M_0.

A decreasing transformation refers to the decrease of certain attributes of the element. In the production process, the reduction of redundant action or work procedures belongs to the decreasing transformation of event-element, which can significantly improve production efficiency.

(c) Expansion/contraction transformations
Expansion transformation: Quantitative expansion transformation is multiple quantitative expansion of a basic-element. As for a matter-element, its quantitative expansion transformation will inevitably lead to expansion transformation of the matter. For example, the volume expansion of a balloon will inevitably lead to expansion of the balloon itself.

Contraction transformation: As for matter-element, its quantitative contraction transformation will inevitably lead to contraction transformation of the matter.

(d) Decomposition/combination transformations
Decomposition transformation refers to division of one object or attributes into several pieces. On the contrary, combination transformation refers to combination of several objects or attributes into a whole one. For example, one action can be executed in several steps.

(e) Duplication transformation
Duplication transformation refers to duplication of the basic-element to multiple basic-elements, such as photo-processing, copying, scanning, printing, optical disc burning, sound recording, video recording, the method of reuse, and reproduction of products, etc. This kind of transformation is extensively applied in the field of information, such as file copying and pasting.

Based on theory of extension set, knowledge from data mining can be mined in a second level by transformation methods, such as substitution transformation, decomposition or combination transformation, and so on. For example, decision trees can mine explainable rules, but it is only a static know-what knowledge and we may not know how to transfer class *bad* to class *good*. To improve such kind of situations, we focus on a new methodology for discovering actionable know-how knowledge based on decision tree rules and extension set theory. It is useful to re-mine rules from data mining so as to obtain actionable knowledge for wise decision making. The transformation knowledge acquiring solution on decision tree rules are practically used to reduce customer churn.[19]

4. Framework and Process of Extension Collaborative Innovation

4.1. *Framework of Extenics-based innovation*

The innovation method based on Big Data and Extenics would take advantage of specific extension methods to generate new innovative ideas or solutions. A framework is given in Fig. 2 and its relevant steps are listed as follows:

Step 1. *Multi-structure data collection*
Collect data related to the innovation goal^G and practical condition^L from data base, expertise, tacit knowledge such as experience and the web, blogs, etc., according to the method presented in Sec. 3.1.

Step 2. *Build basic-element base*
Describe and transform the information into matter-elements, event-elements, or relation-element. Meanwhile, primary knowledge is discovered from data mart by data mining. Then, we save them in the database as a basic-element tree supported by ontology. After this step, we could get a systematic cube of integrated information,[33] according to the method presented in Sec. 3.2.

Step 3. *Obtain all possible solutions by extension transformation*
Taking basic-elements and primary knowledge as input, extension transformation methods as methodology (as shown in Sec. 3.3), we transform the field, the elements, or the criteria of the goals and conditions on basic-elements that are already explored by Step 2. The detailed description to all possible solutions by human–computer interactions based on extension set theory will be presented in Sec. 4.2.

Step 4. *Scoring and evaluation by dependent function*
All the possible solutions are scored by superiority evaluation method based on dependent function quantitatively and expert's experience qualitatively. Then, results are acquired in feasible innovation proposals.

Suppose measuring indicator sets are $MI = \{MI_1, MI_2, \ldots, MI_n\}$, $MI_i = (c_i, V_i)$, $(i = 1, 2, \ldots, n)$, and weight coefficient distribution is

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n).$$

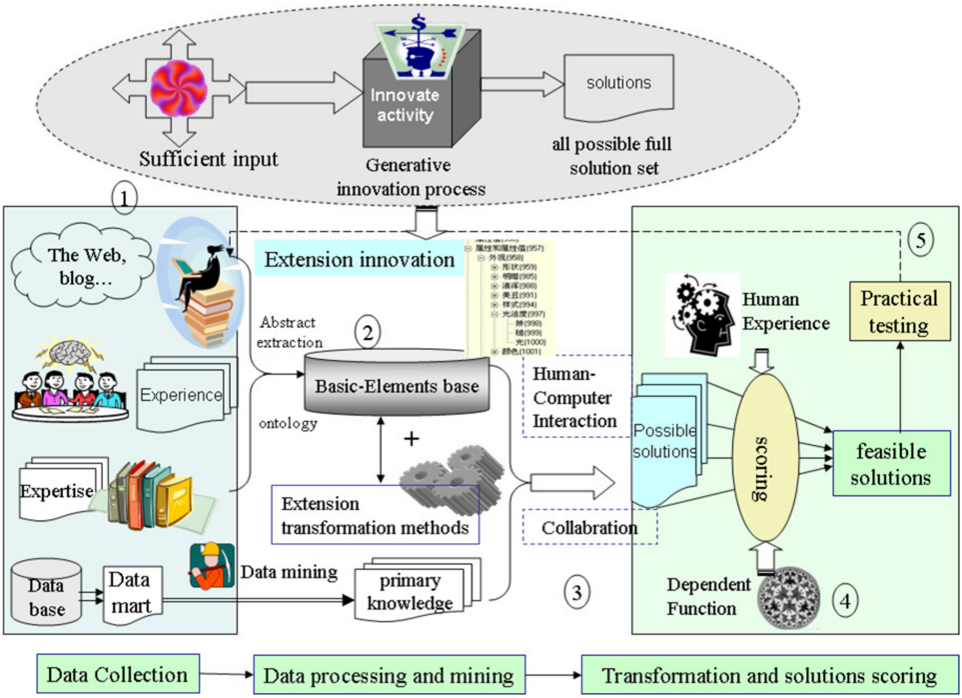


Fig. 2. Framework of extension collabrative innovation model in Big Data.

According to the requirements of every measuring indicator, dependent functions $K_1(x_1), K_2(x_2), \dots, K_n(x_n)$ are established.

The dependent function value of object Z_j about each measuring indicator MI_i is denoted by $K_i(Z_j)$ for easy writing, and the dependent degree of every object $Z_1, Z_2, \dots, Z_m$, about MI_i is

$$K_i = (K_i(Z_1), K_i(Z_2), \dots, K_i(Z_m)), \quad (i = 1, 2, \dots, n).$$

The above dependent degree is standardized as:

$$k_{ij} = \frac{K_i(Z_j)}{\max_{q\in\{1,2,\dots,m\}}|K_i(Z_q)|}, \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m).$$

Then the standard dependent degree of every object $Z_1, Z_2, \dots, Z_m$ about MI_i is

$$k_i = (k_{i1}, k_{i2}, \dots, k_{im}), \quad (i = 1, 2, \dots, n).$$

Step 5. *Practical testing and feedback*
We apply the feasible solutions to practices, collect new data and contrast the results, feedback to the model, and update basic-elements, knowledge base, or methods.

The above given framework of extension innovation model has been used by a reputable company in China and its case study is presented in Sec. 5.

4.2. *Directions of collaborative innovation*

All needed information and knowledge can be described as basic-element, we then take matter (one kind of basic-element) as an example. It has four characteristics and eight aspects as mentioned in Sec. 3.1. There are four main directions for innovation based on matter analysis as shown in Fig. 3.

(1) From the view of descriptions, we can extend and share our ideas from object, character, and its measure; each object has many characters, denoted as $1M : nC$. Similarly, each character can have many values, such as color can be red, green, yellow, or blue. We denote it as $1C : nV$.
(2) From the view of transformation path, we can extend our thinking from goals and conditions to both goals and conditions. Each path has many characters, measures, and objects.
(3) Based on basic extension methods, there are five methods that can be used on the matters as stated in Sec. 3.3.
(4) From the view of transformation objects, we can transform the elements, such as matter, event, or relationships, or transform the criteria and field. For example, a salesman F is regarded as good in company A, but scored as bad after he changed to company B. Here, the salesman F is an element, the rule *good* or *bad* is a criterion, and the companies A and B together are the field.

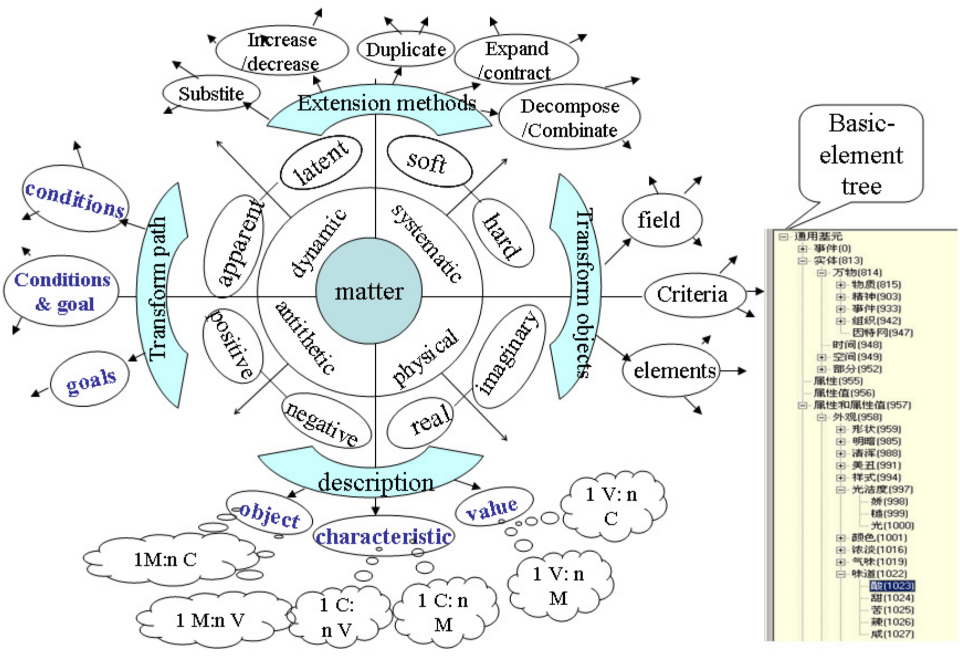


Fig. 3. Four main directions for collabrative innovation.

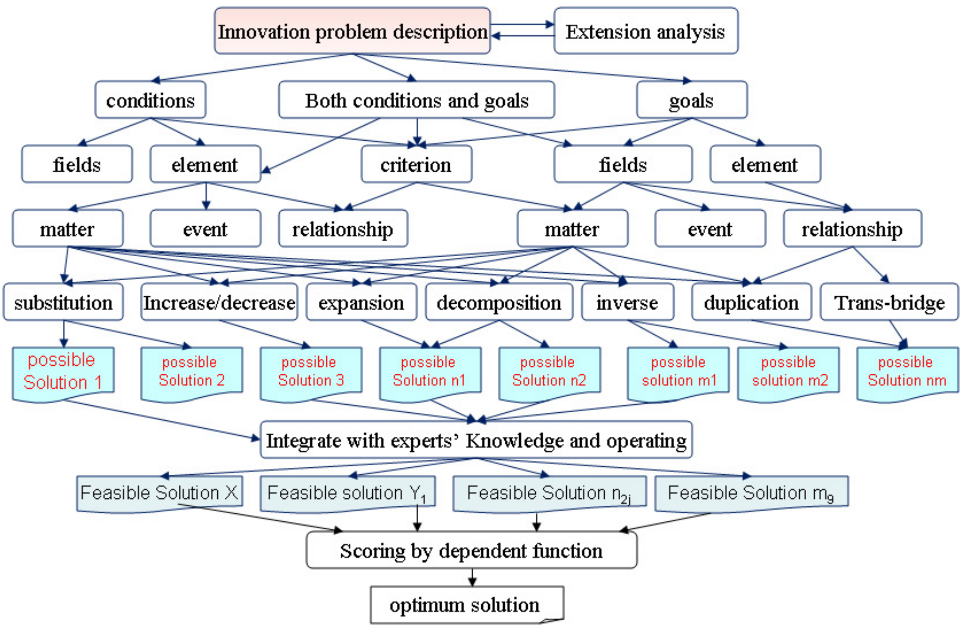


Fig. 4. Map of innovation solutions by extension transformation.

After four main transformations, we can get an information tree both for goals and conditions. Similarly, we can obtain an innovation solution tree. The solution tree is shown in Fig. 4.

4.3. *Scoring method based on Big Data*

Big Data give us a full view to score our innovation solutions. Take 2D as examples according to Extenics.[34]

Extension set theory is a set theory that describes the changing recognition and classification accordingly. Extension set describes the variability of things, using numbers in $(-\infty, +\infty)$ to describe the degrees of how the thing owns certain property, and using an extensible field to describe the reciprocal transformation between the "positive" and "negative" of things. It can describe not only the reciprocal transformation between the positive and negative of things but also the degree of how the thing owns a property.

4.3.1. *Definition of extension set*

Suppose U is universe of discourse, u is any one element in U, k is a mapping of U to the real field I, $T = (T_U, T_k, T_u)$ is a given transformation, we define

$$\tilde{E}(T) = \{(u, y, y') \,|\, u \in U, y = k(u) \in I; T_u u \in T_U U, y' = T_k k(T_u u) \in I\},$$

as an extension set on the universe of discourse U, $y = k(u)$, the dependent function of $\tilde{E}(T)$, and $y' = T_k k(T_u u)$ the extension function of $\tilde{E}(T)$, wherein, T_U, T_k, and T_u are transformations of respective universe of discourse U, dependent function k, and element u (Yang and Cai, 2013).

If $T \neq e$, we define

$\dot{E}_+(T) = \{(u, y, y') \,|\, u \in U, y = k(u) \leq 0; T_u u \in T_U U, y' = T_k k(T_u u) > 0\}$, as *positive extensible field of*$\tilde{E}(T)$; we define

$\dot{E}_-(T) = \{(u, y, y') \,|\, u \in U, y = k(u) \geq 0; T_u u \in T_U U, y' = T_k k(T_u u) < 0\}$, as *negative extensible field* of $\tilde{E}(T)$; we define

$E_+(T) = \{(u, y, y') \,|\, u \in U, y = k(u) > 0; T_u u \in T_U U, y' = T_k k(T_u u) > 0\}$, as *positive stable field* of $\tilde{E}(T)$; we define

$E_-(T) = \{(u, y, y') \,|\, u \in U, y = k(u) < 0; T_u u \in T_U U, y' = T_k k(T_u u) < 0\}$, as *negative stable field* of $\tilde{E}(T)$; and we define

$J_0(T) = \{(u, y, y') \,|\, u \in U, T_u u \in T_U U, y' = T_k k(T_u u) = 0\}$, as *extension boundary* of $\tilde{E}(T)$.

4.3.2. *Dependent function in 1D*

In 1983, Cai Wen defined the 1D-dependent function $K(y)$. Accordingly, if one considers two intervals, X_0 and X, that have no common end point, and $X_0 \subset X$, then:

$$K(y) = \frac{\rho(y, X)}{\rho(y, X) - \rho(y, X_0)}.$$

Since $K(y)$ was constructed in 1D in terms of the extension distance $\rho(.,.)$, we simply generalize it to higher dimensions by replacing $\rho(.,.)$ with the generalized $\rho(.,.)$ in a higher dimension.

4.3.3. *Extension distance in 2D*

Instead of considering a line segment AB as representing an interval $[a, b]$ in R, we consider a rectangle in R^2 enclosing the line such as $AMBN$, where AB is the diagonals of the rectangle. The mid-point of AB is now the center of symmetry O of the rectangle. Let $P(x_0, y_0)$ be a point outside of the rectangle, the coordinates of A be (a_1, a_2) and B be (b_1, b_2), and the point of intersection of the line joining P and O with the rectangle be P' be $(x_{p'}, y_{p'})$. The extension distance in 2D is denoted by $\rho(x_0, y_0)$ and is denoted by

$$\rho((x_0, y_0), AMBM) = |PO| - |P'O| = \pm|PP'|,$$

where $|PO|$ is the distance between P and O, $|P'O|$ is the distance between P' and O, and $|PP'|$ is the distance between P and P' as in coordinate geometry. The mid-point O has coordinates $(\frac{a_1+b_1}{2}, \frac{a_2+b_2}{2})$. Take $x_{P'} = a_1$, now we calculate $y_{P'}$ as

$$y_{P'} = y_0 + \frac{a_2 + b_2 - 2y_0}{a_1 + b_1 - 2x_0}(a_1 - x_0).$$

Therefore P' has the coordinates $P'\Big(x_{P'} = a_1, y_{P'} = y_0 + \frac{a_2+b_2-2y_0}{a_1+b_1-2x_0}(a_1 - x_0)\Big)$.

The distance $d(P,O) = |PO| = \sqrt{(x_0 - \frac{a_1+b_1}{2})^2 + (y_0 - \frac{a_2+b_2}{2})^2}$, while the distance

$$d(P',O) = |P'O| = \sqrt{\left(a_1 - \frac{a_1+b_1}{2}\right)^2 + \left(y_{P'} - \frac{a_2+b_2}{2}\right)^2}$$
$$= \sqrt{\left(\frac{a_1-b_1}{2}\right)^2 + \left(y_{P'} - \frac{a_2+b_2}{2}\right)^2}.$$

Also, the distance $d(P,P') = |PP'| = \sqrt{(a_1-x_0)^2 + (y_{P'}-y_0)^2}$.

Hence the extension 2D-distance formula:

$$\begin{aligned}\rho((x_0,y_0), AMBM) &= P(x_0,y_0), A(a_1,a_2)MB(b_1,b_2)N) = |PO| - |P'O| \\ &= \sqrt{\left(x_0 - \frac{a_1+b_1}{2}\right)^2 + \left(y_0 - \frac{a_2+b_2}{2}\right)^2} \\ &\quad - \sqrt{\left(\frac{a_1-b_1}{2}\right)^2 + \left(y_{P'} - \frac{a_2+b_2}{2}\right)^2} \\ &= \pm|PP'| \\ &= \pm\sqrt{(a_1-x_0)^2(y_{P'}-y_0)^2},\end{aligned}$$

where $y_{P'} = y_0 + \frac{a_2+b_2-2y_0}{a_1+b_1-2x_0}(a_1 - x_0)$.

4.3.4. *Extension distance in n-D*

We can generalize Cai Wen's idea of the extension 1D set to an extension n-D set, and define the *extension n-D distance* between a point $P(x_1, x_2, \ldots, x_n)$ and the n-*D* set S as $\rho((x_1, x_2, \ldots, x_n), S)$ on the linear direction determined by the point P and the optimal point O (the line PO) in the following way:

(1) $\rho((x_1, x_2, \ldots, x_n), S) =$ the *negative distance* between P and the set frontier, if P is inside the set S;
(2) $\rho((x_1, x_2, \ldots, x_n), S) = 0$, if P lies on the frontier of the set S;
(3) $\rho((x_1, x_2, \ldots, x_n), S) =$ the *positive distance* between P and the set frontier, if P is outside the set.

We get the following properties:

(1) It is obvious from the above definition of the extension n-*D* distance between a point P in the universe of discourse and the extension n-*D* set S that:

  (i) Point $P(x_1, x_2, \ldots, x_n) \in \mathrm{Int(S)}$ iff $\rho((x_1, x_2, \ldots, x_n), S) < 0$;
  (ii) Point $P(x_1, x_2, \ldots, x_n) \in \mathrm{Fr(S)}$ iff $\rho((x_1, x_2, \ldots, x_n), S) = 0$;
  (iii) Point $P(x_1, x_2, \ldots, x_n) \notin \mathrm{S}$ iff $\rho((x_1, x_2, \ldots, x_n), S) > 0$.

(2) Let S_1 and S_2 be two extension sets, in the universe of discourse U, such that they have no common end points, and $S_1 \subset S_2$. We assume they have the same optimal points $O_1 \equiv O_2 \equiv O$ located in their center of symmetry. Then, for any point $P(x_1, x_2, \ldots, x_n) \in U$ one has:

$$\rho((x_1, x_2, \ldots, x_n), S_1) \geq \rho((x_1, x_2, \ldots, x_n), S_2).$$

Then we proceed to the generalization of the dependent function from 1D space to n-D space dependent function, using the previous notations.

The dependent n-D function formula is:

$$K_{nD}(x_1, x_2, \ldots, x_n) = \frac{\rho((x_1, x_2, \ldots, x_n), S_2)}{\rho((x_1, x_2, \ldots, x_n), S_2) - \rho((x_1, x_2, \ldots, x_n), S_1)}.$$

5. Case Study

Y group is one of the world's largest menswear manufacturer group, with a production capacity of 80 million clothing items per year, includes shirts, suits, trousers, jackets, leisure coats, knitted items, and ties. Y has implemented world-class modern production lines and high-end equipment from nations including Germany, United States, and Japan. Y group's production and supply lines are using state-of-the-art comprehensive computer-operated technologies to enhance the clothing production process.

After 30 years of development, the group has many subsidiaries, including factories, sales companies, foreign trade corporation, and logistics companies. Y has forged a strong vertically integrated clothing chain, integrating the upstream components of textile and fabric production, midstream component of garment creation, and the downstream components of marketing and sales.

In recent years, the average annual sales increased by 10.1%, costs remained nearly unchanged at the previous year's level. However, the average profit of the group increased only by 0.21%, not significantly improved as shown in Table 3. To find this reason, our data analysis team worked together with the group's financial sector.

The profit of Y group is the sum of its subsidiaries and is denoted as

$$P_g = \sum_{j=1}^{n} P_j = P_f + P_s + P_l + P_t + \cdots, \quad P_g = I_{\text{income}} - C_{\text{cost}}.$$

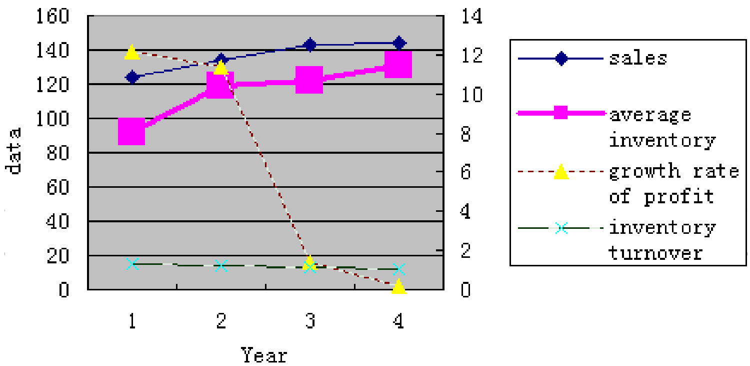
Accordingly, we made a basic-element analysis on profit-related attributes, as shown in Fig. 5.

According to the basic-element analysis, we determine the scope of the collection of Big Data as following:

(1) Total sales, inventory, and profit and cost data of the company for the past five years.

Table 3. Problem seeking on data integratation and analysis.

Items	Year 1	Year 2	Year 3	Year 4
Sales	123.9563	133.6302	143.0152	143.6497
Average inventory	91.6006	118.8934	122.4309	131.1287
Growth rate of profit	12.12	11.38	1.44	0.21
Inventory turnover	1.35	1.2	1.17	1.02



(2) Historical production records of the production plant, raw material records, employee payroll (removal of sensitive personal information), production schedules, storage records, and so on.
(3) Historical distribution data in logistics department, sales returns data, and loss of productions in self-run stores.
(4) Retail sales records in the stores and shops, records of group purchase, discounts records, and inventory balances.

We then integrate data in data warehouse and find information such as which are best-selling products, the highest inventory of suits, inventory turnover days, inventory days of supply and other indicators available as shown in Table 4.

Through analysis of various types of data in the group's departments, we found that:

(1) The main source of profit of the garment sector is self-store sales and group purchase business confirmed by group financial center.
(2) The number of inventory days of supply (similar with inventory turnover) is worse. In some areas, it is up to 536 days, which means that according to the current average sales, inventory of goods available for sale will be 536 days as shown in Fig. 6.
(3) In production scheduling, small orders that are temporarily postponed are as high as 31%, and those into priority processing sequences are mostly original equipment manufacturer (OEM) orders.

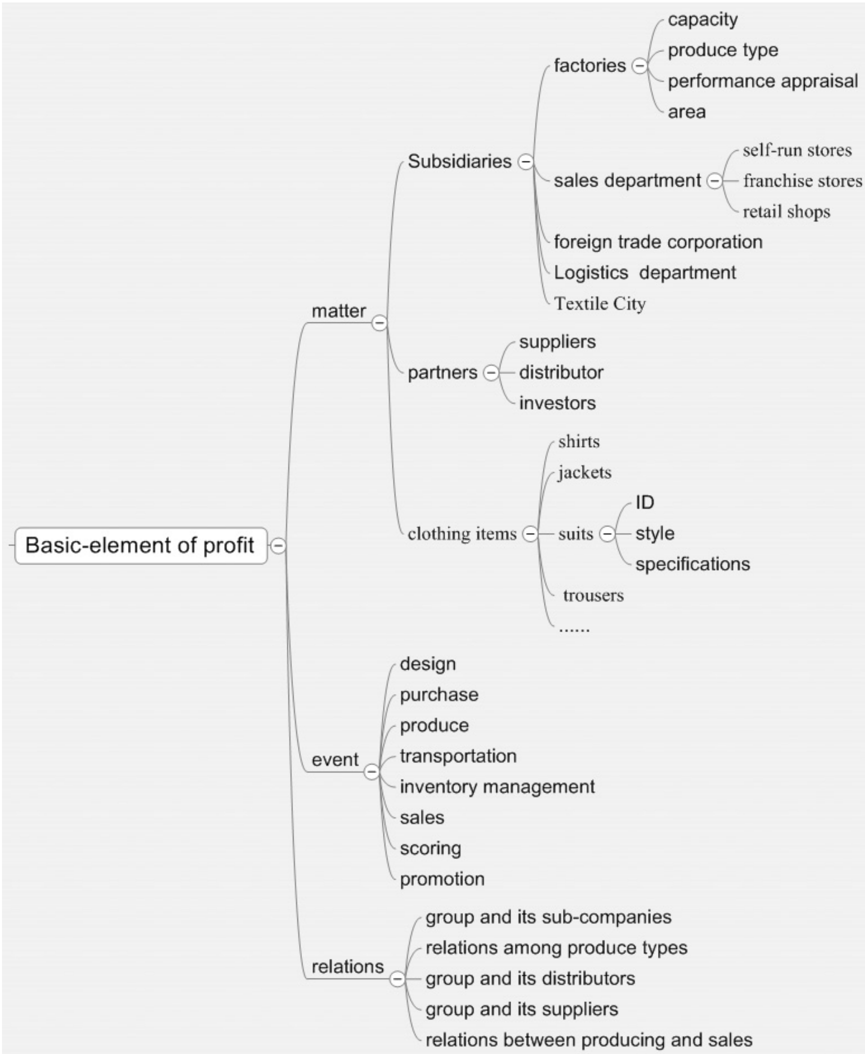


Fig. 5. The basic-element of profit analysis.

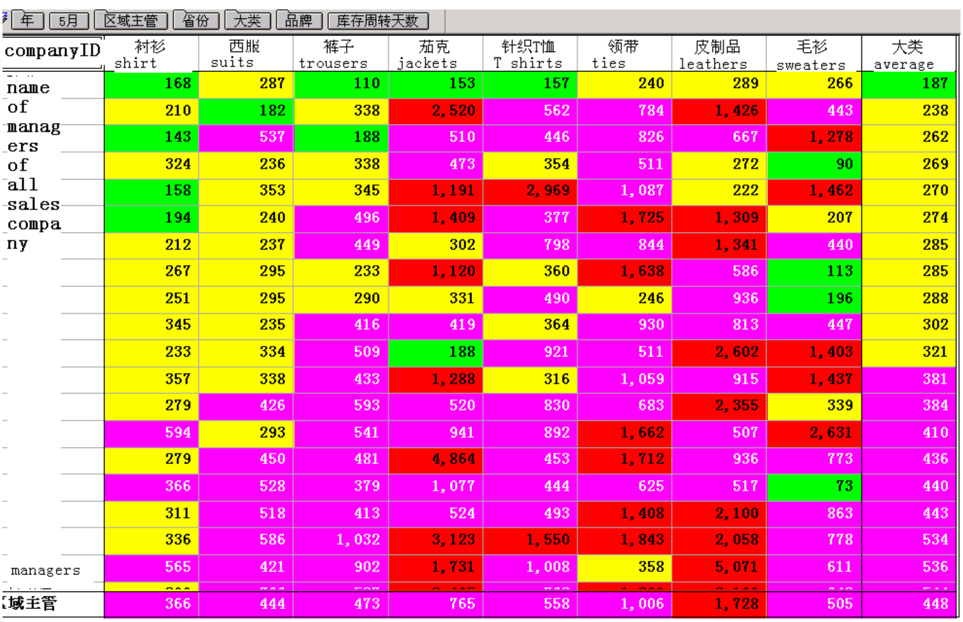
(4) In production plant, the timely delivery rates of group purchase orders have continued to decline, far below than that of the OEM production. OEM orders are usually in large quantities and the proportion in the profits of garment plants is also growing.

Based on extension set theory, we further analyzed the domain and the associated rules. After the analysis of the data, we found that:

(1) Profit is a main key performance indicator (KPI) for the appraisal of subordinate units of the group; in some garment plants annual profit targets are completed

Table 4. Sales analysis and indicators calculation.

Product ID	Sum of sales	Monthly average inventory	Days of inventory turnover	Rate of sales and delivery (%)
Y01V88417	566,566	749,788	40	43.41
Y01T8401A	272,015	272,390	30	48.38
Y01W72052	121,260	212,850	53	34.11
Y01W72059G	118,422	130,806	33	36.43
Y01V88488	110,160	61,560	17	94.44
Y01V88436	108,416	162,118	45	38.23
YC01SCT84396	94,770	42,200	13	99.93
Y01FV88417	84,798	96,788	34	50.76
Y01CA100744	83,448	104,310	38	48.09
Y01CA886112	81,404	53,298	20	87.83
Y01W72050	70,176	146,415	63	23.82
Y01W72053	67,338	192,210	86	23.43
YC01CW72048	67,250	103,125	46	31.72
YC01CW72045	65,750	168,875	77	27.80
Y01V5000	61,600	27,425	13	82.74
Y01T48506G	57,986	136,196	70	32.36
Y01CA88690	56,588	95,504	51	36.80
YC01CW72041	55,250	117,625	64	29.00
Y01T50	54,300	51,525	28	56.59
Y01CA88688	53,016	94,376	53	34.52
Y01T48505G	52,140	83,108	48	42.42
YC01CW72044	52,000	166,000	96	21.16



companyID	衬衫 shirt	西服 suits	裤子 trousers	茄克 jackets	针织T恤 T shirts	领带 ties	皮制品 leathers	毛衫 sweaters	大类 average
	168	287	110	153	157	240	289	266	187
	210	182	338	2, 520	562	784	1, 426	443	238
	143	537	188	510	446	826	667	1, 278	262
	324	236	338	473	354	511	272	90	269
	158	353	345	1, 191	2, 969	1, 087	222	1, 462	270
	194	240	496	1, 409	377	1, 725	1, 309	207	274
	212	237	449	302	798	844	1, 341	440	285
	267	295	233	1, 120	360	1, 638	586	113	285
	251	295	290	331	490	246	936	196	288
	345	235	416	419	364	930	813	447	302
	233	334	509	188	921	511	2, 602	1, 403	321
	357	338	433	1, 288	316	1, 059	915	1, 437	381
	279	426	593	520	830	683	2, 355	339	384
	594	293	541	941	892	1, 662	507	2, 631	410
	279	450	481	4, 864	453	1, 712	936	773	436
	366	528	379	1, 077	444	625	517	73	440
	311	518	413	524	493	1, 408	2, 100	863	443
	336	586	1, 032	3, 123	1, 550	1, 843	2, 058	778	534
managers	565	421	902	1, 731	1, 008	358	5, 071	611	536
域主管	366	444	473	765	558	1, 006	1, 728	505	448

Fig. 6. Inventory analysis by sales.

more than 95% and the average annual profit growth is of 3.87% — one of the best performance units.

(2) In some sales companies, sales increase at an average rate of 9.01%, while the inventory grows at a rate of 14.2%, and whereas the profit growth rate is only 0.53%.

Based on basic-element theory of Extenics, we have collected attributes information related to garment plants:

(1) The produce types are divided into three categories:
   - Make-to-stock production, for the production of inventory for all kinds of self-run stores sales;
   - Make-to-order production, especially for group purchase;
   - OEM. The plants manufacture clothing that are purchased by another international company and retailed under that purchasing company's brand name.

(2) The OEM processing fee is on average 2–4 Yuan higher than the make-to-order processing fee.

(3) By deep inquiry with managers, we found that in the production plant, the processing priority rules are: high profit, large quantities of orders will have priority processing, and OEM orders meet these two conditions.

Integrating the knowledge mined from the Big Data and information basic elements, we draw a knowledge logic chain as follows:

Profit is the primary KPIs of production plants, therefore managers give the top priority to process the high-profit orders, and since OEM orders are latent with high profits, they process OEM orders by priority, and since OEM orders are large and production capacity is limited, make-to-order productions are postponed.

It is reasonable to process high-profits OEM orders from the perspective of the garment plants, but it is not understandable from the group's perspective; OEM processing fee is so small that it is almost negligible compared to the profit of a group buy ($\leq 5\%$); to earn \$4, they lost \$80.00. One of the reasons for the cause of this "penny wise, pound foolish" phenomenon is the improper setup of profit as KPI for garment plants.

By Big Data analysis and extension model, we selected and added two new KPIs defined as *ratio of timely delivery of order* and *operation costs* to replace with single KPI profit. Therefore, we score the garment plants by 3D-dependent functions. If we only use profit as KPI, SH plants ranked A; however, if we use three KPIs and score with 3D-dependent functions, SH will be ranked C. This greatly helped the managers: the production capacity has been fully utilized in garment plants, timely delivery rate of group-buy-order is greatly improved, and OEM orders delivery rate is almost keep the same, while the group's overall profit improved from 0.21% to 12.6%.

Last year, Y group's total assets are valued at 30 billion RMB. Its annual textile and garment sales alone amount to 10 billion RMB. Y was ranked number 113 in China's list of its top 500 manufacturers and also is the only clothing company that China has recognized as one of its "Advanced Manufacturers."

6. Summary

The paper presents a framework of extension collaborative innovation model in the context of Big Data with concrete processes and a case study. The model integrates Extenics, data mining, and knowledge management, and develops a framework for collaborative innovation with team work. By collecting knowledge or information from multiple resources among all departments, we can build information tree in basic elements from various forms of data. Knowledge or information related to the problem can find relations by human–computer interaction method.

This particular method combines qualitative analysis which would take advantage of personal intelligence after formalized expression of innovation problems and quantitative analysis which follows a systematic flow based on accumulated knowledge or innovation patterns. It helps to solve management problems according to the extensibility of basic element and was applied in the innovation of management beyond data technology such as data mining and intelligent knowledge management.

By case study, we found that Extenics can serve as the starting point of a generative approach for collaborative innovation, because it focuses on solving non-compatible problems by formularized methods (Yang and Cai, 2013). The main features of using Extenics for collaborative innovation in big data can be outlined as follows:

- Extenics provides a system structure for data collection and processing. Then, it offers an opportunity for generating all possible solutions for innovative problem solving.
- In addition to generating possible solutions, Extenics also offers an effective way of evaluating solutions, so that any solutions failing the evaluation will be eliminated from further consideration. This can prevent computational explosion.
- The Extenics-based approach for innovation is a human–computer interactive process: Computers can conduct scalable data storage and mining by making use of various algorithms, far superior than what human labors can handle. Nevertheless, the real world is extremely complex and only humans can capture the dynamics of the real world beyond what any algorithms can handle.

From these general features, we provided an overview on what Extenics can offer for collaborative innovation. First, we had put forward a data collection theory in Sec. 3 to acquire sufficient input from Big Data to generate ideas, from which innovation process can take place in a systematic way, instead of by chance. Second, we presented a combined model to process Big Data — one way is to build basic-element tree by human, another way is to build knowledge base by computer such as data mining, extension transformation connect and help them to generate possible solutions in all directions by a formalized method. Last, we extend the dependent function to score multi-attribute innovation solutions in the context of Big Data. By this model, we can obtain novel ideas from several ordinary rules mined from multi-data source.

In the future, we will further test the n-D-dependent function and compare it with other methods. Moreover, the basic-element base and knowledge collaborative methods need to be integrated with agent-based system and enhance the extension innovation model. Due to the significant importance of Big Data, deep research about combination of above methods with web IT, extension data mining, and intelligent knowledge management need to be further explored. How to update basic-element base automatically and simulate the knowledge innovation process by intelligent agent is a challenging problem.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (#71271191, #11271361 and #70871111), Zhejiang Philosophy and Social Science Project (#11JCSH03YB), the Ministry of Water Resources Special Funds for Scientific research on Public Causes (No. 201301094) and Ningbo Soft Science Research Program (#2011F10008), China.

References

1. E. Hippel and R. Katz, Shifting innovation to users via toolkits, *Management Science* **48**(7) (2002) 821–833.
2. A. Cantisani, Technological innovation processes revisited, *Technovation* **26** (2006) 1294–1301.
3. J. Birkinshaw, G. Hamel and M. J. Mol, Management innovation, *Academy of Management Review* **33**(4) (2008) 825–845.
4. B. Chae and D. L. Olson, Business analytics for supply chain: A dynamic-capabilities framework, *International Journal of Information Technology and Decision Making* **12**(1) (2013) 9–26.
5. M. Zeleny, High technology and barriers to innovation: From globalization to relocalization, *International Journal of Information Technology and Decision Making* **11**(2) (2012) 441–456.
6. J. Bughin, M. Chui and J. Manyika, Clouds, big data, and smart assets: Ten tech-enabled business trends to watch, *McKinsey Quarterly* **56** (2010) 1–14.
7. M. Chisholm, Big data and the coming conceptual model revolution (2012), Available at immagic.com.
8. J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers, Big data: The next frontier for innovation, competition, and productivity, Report, McKinsey Global Institute (2011), pp. 1–137.
9. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd edn. (Morgan Kaufman Publishers, San Francisco, CA, 2006).

10. D. Olson and Y. Shi, *Introduction to Business Data Mining*, International edition (McGraw-Hill, 2007).
11. C. Ji, Y. Li, W. Qiu, Y. Jin, Y. Xu, U. Awada, K. Li and W. Qu, Big data processing: Big challenges and opportunities, *Journal of Interconnection Networks* **13**(3–4) (2012) 1250009, 1–19.
12. X. Li, Y. Shi and L. Zhang, *From the Information Explosion to Intelligent Knowledge Management* (Science Press, Beijing, 2010).
13. Y. Shi and X. Li, Knowledge management platforms and intelligent knowledge beyond data mining, in *Advances in Multiple Criteria Decision Making and Human Systems Management: Knowledge and Wisdom*, eds. Y. Shi, D. L. Olsen and A. Stam (IOS Press, Amsterdam, 2007), pp. 272–288.
14. L. Zhang, J. Li, Y. Shi and X. Liu, Foundations of intelligent knowledge management, *Human System Management* **28**(4) (2009) 145–161.
15. W. Cai, Extension set and non-compatible problems, in *Advances in Applied Mathematics and Mechanics in China* (International Academic Publishers, Peking, 1983), pp. 1–21.
16. C. Yang and W. Cai, *Extenics: Theory, Method and Application* (Science Press, Beijing, 2013).
17. P.-L. Yu, *Habitual Domains* (Springer-Verlag, Kansas, KS, 1995).
18. T. L. Saaty, *Creative Thinking, Problem Solving & Decision Making*, 4th edn. (RWS Publications, Pittsburgh, PA, 2010).
19. X. Li, H. Zhang, Z. Zhu, Z. Xiang, Z. Chen and Y. Shi, An intelligent transformation knowledge mining method based on Extenics, *Journal of Internet Technology* **14**(2) (2013) 315–325.
20. B. A. Hennessey and T. M. Amabile, Creativity, *Annual Review of Psychology* **61** (2010) 569–598.
21. M. D. Mumford, K. E. Medeiros and P. J. Partlow, Innovation: Processes, strategies, and knowledge, *Journal of Creative Behavior* **46**(1) (2012) 30–47.
22. G. Kou, Y. Lu, Y. Peng and Y. Shi, Evaluation of classification algorithms using MCDM and rank correlation, *International Journal of Information Technology and Decision Making* **11**(1) (2012) 197–225.
23. M. Sorli and D. Stokic, Future trends in product/process innovation, *International Journal of Innovation and Technology Management* **8**(4) (2011) 577–599.
24. A. H. Vandeven, Central problems in the management of innovation, *Management Science* **32**(5) (1986) 590–607.
25. Z. Hua, J. Yang, S. Coulibaly and B. Zhang, Integration TRIZ with problem-solving tools: A literature review from 1995 to 2006, *International Journal of Business Innovation and Research* **1**(1–2) (2006) 111–128.
26. Z. Zhou and X. Li, Research on Extenics-based innovation model construction and application of enterprise independent innovation, *Studies in Science of Science* **28**(5) (2010) 769–776.
27. K. A. Ericsson and J. H. Moxley, The expert performance approach and deliberate practice: Some potential implications for studying creative performance in organizations, in *Handbook of Organizational Creativity*, Ed. Michael D. Mumford (Academic Press, Salt Lake City, 2012), pp. 141–167.
28. D. Van Horn, A. Olewnik and K. Lewis, Design analytics: Capturing, understanding, and meeting customer needs using big data, *ASME Int. Design Engineering Technical Conf. Computers and Information in Engineering Conf.* (Chicago, Illinois, USA, 2012) 1–13.

29. G. Kou and C. Lou, Multiple factor hierarchical clustering algorithm for large scale web page and search engine clickstream data, *Annals of Operations Research* **197**(1) (2012) 123–134.
30. C. Snijders, U. Matzat and U.-D. Reips, "Big Data": Big gaps of knowledge in the field of internet science, *International Journal of Internet Science* **7** (2012) 1–5.
31. M. Hilbert, Big data for development: From information to knowledge societies (2013), Available at SSRN: http://ssrn.com/abstract=2205145 (accessed on 24 January 2013).
32. J. Kim, A. Lund and C. Dombrowski, Telling the story in big data, *Interactions* **20**(3) (2013) 48–51.
33. X. Li, H. Qu, Z. Zhu *et al.*, A systematic information collection method for business intelligence, *Int. Conf. ECBI* (2009), pp. 116–119.
34. F. Smarandache, Generalizations of the distance and dependent function in Extenics to 2D, 3D, and n-D, *Global Journal of Science Frontier Research* **12**(8) (2012) 47–60.