Chance Discovery: The Current States of Art

Yukio Ohsawa

School of Engineering, The University of Tokyo ohsawa@q.t.u-tokyo.ac.jp

1 Introduction: Scenarios and Chances in the Basis of Decisions

As defined in [18],a *chance* in chance discovery means to understand an unnoticed event/situation which can be (uncertain, but) significant for making a decision. Associating the event with the appearance of a product or a service, customers seek a valuable event, i.e., the appearance of a product/service significant for his/her decision to improve daily life. Associating the event with a message from a customer, people in the side of business should look at a valuable event, i.e., a message significant for the decision to improve the service.

Chance discovery is an essential basic research area applicable to all kinds of business. The above definition of a chance may have sounded counter-intuitive for reader thinking about an accident or uncertainty, say events occurring *by chance*. To such an opinion, we have been asserting chance discovery means the discovery *of* chance, not *by* chance. However, according to the recent progress of studies on the methods of chance discovery, the relevance between discovery *of* chance and discovery *by* chance came to be more positively recognized. That is, a chance defined as an event significant for decision making has all the natures of a chance in the phrase "by chance," i.e. (1) uncertainty, (2) accident (3) probability, if we introduce scenario-based thoughts about chance discovery, and these can all be put into a power of survival.

Note here, that a decision means to choose one from multiple possible scenarios of future events and actions, so there is "uncertainty" in (1), in the future scenarios where chance discovery is desired. In other words, uncertainty can be the motivation to decide something. Therefore, "probability" in (3) rather than True/False may become an appropriate measure of the justification of a scenario. An "accident" of (2) implies uncertainty to lead to an opportunity or to a risk, relying on the future scenario.

In general, a scenario is a time series of events/states to occur under a certain context. And, a chance can be regarded as the cross of multiple scenarios.

Y. Ohsawa: *Chance Discovery: The Current States of Art*, Studies in Computational Intelligence (SCI) **30**, 3–20 (2006)
www.springerlink.com

For example, suppose a customer of a drug store buys a number of items in series, a few items per month. He should do this because he has a certain persistent disease. In this case, a remedy of the disease suggested from his doctor is the context shared over the event-sequence, where an event is this patient's purchase of a set of drugs. This event-sequence is a scenario under the context proposed by the doctor. However, the patient may hear about a new drug, and begin to buy it to change the context, from the remedy he followed so far to a new remedy to seep up his cure. In other words, the patient introduces a new scenario. After a month, his doctor gets upset hearing this change due to the patient's ignorance about the risk of the new drug. The doctor urgently introduces a powerful method to overcome all the by-effects of the risky new drug - changing to the third scenario.

In this example, we find two "chances" in the three scenarios. The first chance is the information about the new drug which changed from the first remedy scenario to the second, riskier one. Then the doctor's surprise came to be the second chance to turn to the third scenario. Under the definition of "chance," i.e., an event or a situation significant for decision making, a chance occurs at the cross point of multiple scenarios as in the example because a decision is to select one scenario in the future. Generally speaking, a set of alternative scenarios form a basis of decision making, in domains where the choice of a sequence of events affects the future significantly. Based on this idea, the methods of chance discovery have been making successful contributions to science and business domains [1].

Now let us stand on the position of a surgeon looking at the time series of symptoms during the progress of an individual patient's disease. The surgeon should make appropriate actions for curing this patient, at appropriate times. If he does so, the patient's disease may be cured. However, otherwise the patient's health condition might be worsened radically. The problem here can be described as choosing one from multiple scenarios. For example, suppose states 4 and 5 in Eq. (1) mean two opposite situations.

$$\begin{aligned}\text{Scenario 1} &= \{\text{state } 0- > \text{state } 1- > \text{state } 2- > \text{state } 3- > \\ &\qquad \text{state } 4 \text{ (a normal condition)}\}. \\ \text{Scenario 2} &= \{\text{state } 4- > \text{state } 5- > \text{state } 6 \text{ (a fatal condition)}\}. \end{aligned} \tag{1}$$

Each event-sequence in Eq. (1) is called a *scenario* if the events in it share some common context. For example, Scenario 1 is a scenario in the context of cure, and Scenario 2 is a scenario in the context of disease progress. Here, suppose there is a hidden state 11, which may come shortly after or before state 2 and state 5. The surgeon should choose an effective action at the time of state 2, in order to turn this patient to state 3 and state 4 rather than to state 5, if possible. Such a state as state 2, essential for making a decision, is a *chance* in this case.

Detecting an event at a crossover point among multiple scenarios, as state 2 above, and selecting the scenario going through such a cross point means a chance discovery. In general, the meaning of a scenario with an explanatory context is easier to understand than an event shown alone. In the example of the two scenarios above, the scenario leading to cure is apparently better than the other scenario leading to a fatal condition. However, the meaning of chance events, which occurs on the bridge from a normal scenario to a fatal scenario, i.e., state 2, state 11, and state 5 in Fig. 1, are hard to understand if they are shown independently of more familiar events. For example, if you are a doctor and find polyp is in a patient's stomach, it would be hard to decide to cut it away or to do nothing else than leaving it at the current position. On the other hand, suppose you find the patient is at the turning point of two scenarios - in one, the polyp will turn larger and gets worsened. In the other, the polyp will be cut away and the patient will be cured. Having such scenarios, you can easily choose the latter choice.

Consequently, an event should be regarded as a valuable chance if the difference of the merits of scenarios including the event is large, and this difference is a measure of the utility of the chance. Discovering a chance and taking it into consideration is required for making useful scenarios, and proposing a number of scenarios even if some are useless is desired in advance for realizing chance discovery. For realizing these understandings, visualizing the scenario map showing the relations between states as in Fig. 1 is expected to be useful. Here, let us call each familiar scenario, such as Scenario 1 or Scenario 2 , an *island*. And, let us call the link between islands a *bridge*. In chance discovery, the problem is to have the user obtain bridges between islands, in order

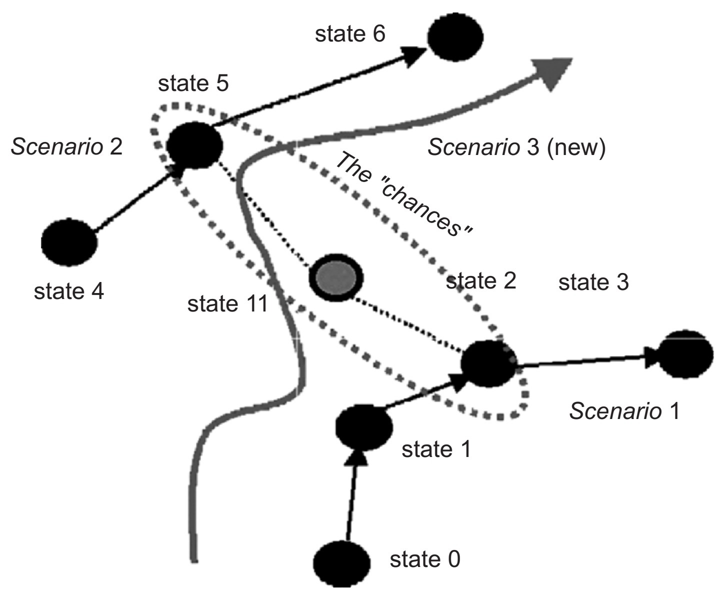

Fig. 1. A chance existing at the cross point of scenarios. The scenario in the thick arrows emerged from Scenario 1 and Scenario 2.

to explain the meaning of the connections among islands via bridges, as a scenario expressed in understandable language.

The research goal of chance discovery has been to enable to choose useful scenarios at the time one should do so, i.e., at the time of a chance, and to accelerate the decision to act on the optimal scenario.

2 Scenario "Emergence" in the Mind of Experts

In the term "scenario development", a scenario may sound like something to be "developed" by human(s) who consciously rules the process of making a scenario. However, valuable scenarios really "emerge" by unconscious interaction of humans and their environment. This occurs like the event 2 appears for itself and all the events in Fig. 1 self-organizes a complex new scenario as the think curved arrow.

For example, a *scenario workshop* developed by the Danish Board of Technology [26] starts from scenarios preset by writers, then experts in the domain relevant to the preset scenarios discuss to improve the scenarios. The discussants write down their opinions during the workshop, but it is rare they notice all the reasons why those opinions came out and why the scenarios have got obtained finally. Rather, new ideas emerge by the self-organizing interactions of particles of ideas from participants. As Montero writes in Chapter 3, self-organization is in the essence of scenario emergence.

This process of scenario workshop can be compared with the KJ (Kawakita Jiro) method, the method in the origin of creation aid, where participants write down their initial ideas on KJ cards and arrange the cards in a 2D-space in co-working for finding good plans. Here, the idea on each card reflects the future scenario in a participants' mind. The new combination of proposed scenarios, made during the arrangement and the rearrangements of KJ cards, helps the emergence of new valuable scenarios, putting in our terminology. In some design processes, on the other hand, it has been pointed out that ambiguous information can trigger creation [6]. The common points among the scenario "workshop", the "combination" of ideas in KJ method, and the "ambiguity" of the information to a designer is that scenarios presented from the viewpoint of each participant's environment are bridged via ambiguous pieces of information about different mental worlds they attend. From these bridges, each participant recognizes situations or events which may work as "chances" to import others' scenarios to get combined with one's own. This can be extended to other domains than designing. In the example of Eq. (1), a surgeon who almost gave up because he guessed his patient is in Scenario 2, may obtain a new hope in Scenario 1 proposed by his colleague who noticed that state 2 is bridging to both scenarios - only if it is still before or at the time of state 2. Here, state 2 is uncertain in that its future can potentially go in two directions, and this uncertainty can make a chance, an opportunity not only a risk.

3 The Process of Chance Discovery

Note that the difference of chance discovery and data mining is that chance discovery aims at obtaining the chances at the cross points of meaningful scenarios, whereas data mining obtains meaningful patterns in the data. If a method of data mining can obtain such pattern as “(state 1 -> state 2) -> state 11 -> (state 5 -> state 6)” in Fig. 1, where the cross of Scenario 1 and Scenario 2 is included, then we can say such a tool can be *used for* chance discovery. However, in general, a data mining tool which can obtain such a complex pattern tends to obtain a huge number of other long patterns, and finally human (user) should choose the most meaningful pattern. As a result, in chance discovery, a critical thing to consider is human’s thoughts for choosing meaningful scenarios, and a data mining tool can be a powerful support for the thinking human(s).

If co-workers in a company section are discussing in order to choose a future scenario of business, their empathy with a proposed scenario is necessary for taking the scenario into their new actions. This empathy leads to not only logical thinking and agreement, but also the mutual understanding of each other’s stand points from the level of emotion. By the coupling of scenarios under participants’ empathy with scenarios and the underlying daily lives of proposers, a scenario meaningful for the team working tends to emerge. We call this a scenario emergence in communication. And, a data mining tool to be mentioned hereafter which has a function to visualize a scenario map as in Fig. 1 can support the coupling, the creation, and the choice of scenarios. As well, tools for predicting rare events [4, 8, 9] plays a significant role in the process of chance discovery.

Suppose we have two scenarios, Scenario 1 and Scenario 2 in Fig. 1, a new scenario may emerge as in the thick arrows. Here, the scenarios make a crossover and generate a new one, like the crossover of chromosomes generating a new chromosome. It is easy to write this way. However, we should stop here and think carefully. We should imagine a simple but real setting in business to consider the difficulties to communicate having empathy with the proposed scenarios and with the underlying daily lives of participants. Let us take an example of a chain of sushi-bars. Traditional sushi-bars were not big companies. However, the recent outbreak of low-price sushi-bars made a number of sushi-bar chain companies. A chain may have tens of sushi-bars all over Japan. In this situation, a chain involves sushi-masters, the head of each sushi-bar, central managers controlling a number of sushi-bars, advertisement and customer-relationship sections, etc. From these staffs, let us pick (a) 10 sushi masters from two local sushi-bars (five from each), and (b) five members of the customer-relation section of the company. The problem is whether or not these 15 staffs can find a coherent decision, i.e., consensus, in developing a new sushi item.

Sushi-master 1 in bar A) I often see customers eating O-toro and Chu-toro. Both items are of oily meat of tuna, so I think we should make original oily tuna items.

Sushi-master 2 in bar B) Your customers are rich, because your place is in a high-class area Ginza. Our customers are students, so we have to increase squids and sardines.

Sushi-master 3 in bar B) No, no! Students take O-toro if we serve in our price.

Sushi-master 1 in bar A) Why don't you think of Ise-ebi (king robster)? This is still rare, but the customers look very happy when they eat one.

Sushi-master 2 in bar B) Umm··· You do not know students ··· Students can not even imagine the taste of Ise-ebi. How can they order Ise-ebi when they can not imagine its taste?

Customers relation staff 1) Why don't you propose the customers to try Ise-ebi with any explanation about its taste, if you think they do not know it?

Sushi-master 3 in bar B) Students are very greedy ··· We are too busy making what they order. How can we explain about Ise-ebi?

Customers relation staff 2) We may have to make a pamphlet of new items, including Ise-ebi, to inform customers about their tastes.

Sushi-master 3 in bar B) No! You cannot move their interest with such papers! They do not read before eating!

As found here, a difficulty lies in finding a consensus in scenario communications. The reason is that the participants have different background domains, and it is hard to present in advance the details of all expertise to be used in their own thoughts for communication. For this reason, they have to discover possible points of consensus, with considering the overview of each other's background domain. And, each domain is too complex to be understand quickly in the beginning of communication. They can not even ask suitable questions about each other's experiences.

Recently, scenario communication came to be supported by tools and theories on chance discovery. The communication is positioned in the process of chance discovery as in Fig. 2, which illustrates the process of chance discovery, called the Double Helix (DH) model [19, 21] . This process starts from a state of user's mind concerned with a new chance, and this *concern* is reflected to collecting *external data*, i.e., data from the object environment, to be visualized by a data-mining tool such as KeyGraph and IDM introduced in the next section, specifically designed for chance discovery. By exploring on this map, i.e., the visual result of data-mining, basic scenarios and their values rise in each user's mind. Then users join a workshop for chance discovery sharing the map. In this workshop, participants begin to understand the relations of *islands* (familiar contexts for participants) and *bridges* (unfamiliar relations between islands) in the map.

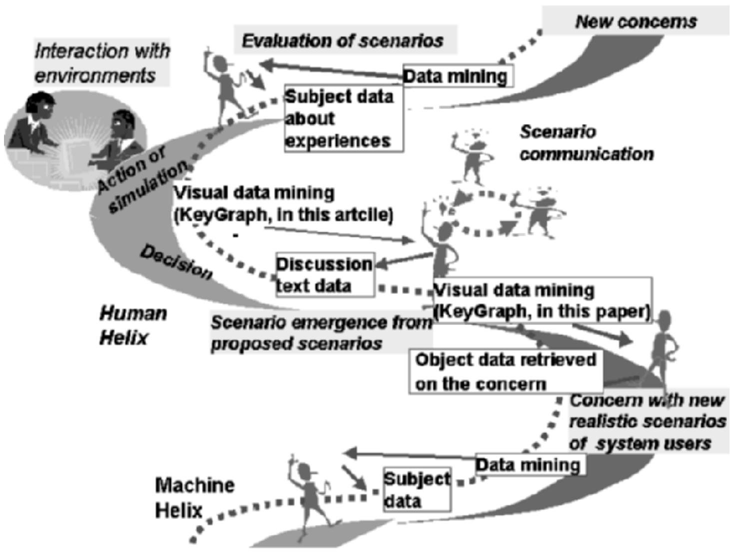

Fig. 2. The double helix process of chance discovery [19].

In the next step, a visual data mining has been applied to the *internal data*, i.e., the data recording the thoughts and messages of participants. Looking at this new map of messages during the communication, the participants notice their own awareness of bridges between basic scenarios, and the islands get connected to have novel scenarios emerge. Here participants may discover chances, because each visualized island corresponds to a basic scenario familiar to some participants and a bridge means a cross of those basic scenarios. That is, connecting the islands via bridges generates a scenario, not confined to an existing island and not too unfamiliar to understand. Based on the scenario selected from those generated here, participant(s) make or simulate actions, and obtain concerns with newer chances - the spiral process progresses to the initial step of the next cycle. As a matter of fact, in the number of successful cases chance discovery were realized by following the steps of double helix [1] (See Chapter by Ohsawa and Usui in Part V).

Even if we get a bunch of visualization tools, the double helix may take the inefficient iteration of cycles, in the worst case forcing to take months for one chance. If the empathy among participants is hard to establish during these cycles, this long process may lead only to a flood of meaningless conflictions. Even worse, the complexity of diagrams visualizing the scenario maps sometimes disturbs communications. Thus, as well as tools for data analysis and visualization, environments for communications and thoughts have been developed for chance discovery.

4 Tools for Chance Discovery and the Environment for Scenario Communications

In this section, let me introduce some of our technical developments. These are formed of tools for visualizing scenario maps and the environment for collaborators' communication about future scenarios.

4.1 Tools for Chance Discovery

We show briefly the outlines of two tools, which have been the most frequently used in real projects of chance discovery in companies. The details of computation schema are presented in the following chapters, and some chapters further improves these methods for respective purposes.

KeyGraph [19, 25] A map showing the relations of events/states in the target environment is visualized. For example, a map of the market is shown on the metaphor of islands (established clusters of consumers) and bridges (unnoticed relations between consumers in islands), if it is applied to consumption data [17]. When it is applied to the internal data, i.e., messages in communication, islands show basic opinions and bridges show new ideas connecting basic opinions. In a textile company, KeyGraphs having real pieces of textile products put on its surface has been introduced. This realized an exceptional growth in their sales performance [30](Chapter 19 of this book), and made a trigger to start their process of chance discovery. In DISCUS, the collaboration of Illinois University and NCSA, KeyGraph is made from communication content, and is always visualized during the on-line communication. This enabled to enhance innovative scenario-communications in the domain of marketing. The recent versions of KeyGraph [7] (Chapter 20 of this book) have the functions to accept user's annotations on the graphical results.

A visualized map showing the relations of events/states in the target environment is useful for participants of a scenario communication in exploring event-relations where scenarios can be drawn, as in Fig. 1, based on personal experiences in their minds. We call this map a scenario map. KeyGraph is a tool for visualizing a scenario map. If the environment represents the world of daily life of people, an event (e.g., "Q3-1") may represent an answer (e.g., choosing '1' from '1' and '0') to a question (e.g., "Q3") in a questionnaire about daily lifestyle. By visualizing the map where answers appear in a graph as in Fig. 3, one can see the overview of the behaviors of survey subjects. In this case, a period ('.') is put at each end of one subject's answer-set. E.g, let $D1$ be: $D1 = {}''Mr.\ A$: Q1-1 Q2-1 Q3-1.

$$\begin{aligned}
&Mrs.B : \text{Q1-1 Q2-1 Q3-1 Q4-1.}\\
&Mr.C : \text{Q4-1 Q5-1 Q7-1 Q8-1.}\\
&Mrs.D : \text{Q5-1 Q2-1 Q3-1 Q5-1 Q7-1 Q8-1.}\\
&Ms.E : \text{Q1-1 Q2-1 Q7-1 Q8-1 Q9-1.}\\
&Mr.F : \text{Q5-1 Q7-1 Q8-1 Q9-1.}
\end{aligned} \tag{2}$$

KeyGraph, of the following steps, is applied to D1 ([20]for details).

KeyGraph-Step 1: Events appearing many times in the data (e.g., "Q1-1" in Eq. (2)) are depicted with black nodes, and each pair of such frequent events occurring often in the same set (e.g., in the same sequence ending with a period) is linked via solid lines. For example, Q1-1, Q2-1, and Q3-1 from Eq. (2) are all connected with a solid line. Each connected graph obtained here forms one island, implying a basic context underlying the belonging events. A clustering method as in [13] can be applied here.

KeyGraph-Step 2: Events which may not be so frequent as the black nodes in islands but co-occurring with multiple islands, e.g., "Q4-1" in Eq. (2), are obtained as hubs. A path of links connecting islands via hubs is called a bridge. If a hub is rarer than black nodes, it is colored in a different color (e.g. red) than black. We can regard such a new hub as a candidate of chance, i.e., an event significant for context-jumping decisions.

In the example of Fig. 3, the result of KeyGraph, the island {Q1-1, Q2-1, Q3-1} means the basic context of established popularity e.g. preference to use mobile phones, and the island of {Q5-1, Q7-1, Q8-1} shows another basic context such as the preference to listen to music by CD players. Then, the bridge "Q4-1" representing an answer "Yes, I use a mobile phone for listening to new music" may mean the instrument for listening to music can change from CD players to mobile phones. If there are clues to determine temporal or causal directions between events, the user may put arrows to the links in the corresponding directions. Then, the scenario map can be the basis for drawing scenarios.

In Fig. 4, the result of KeyGraph for D2 in Eq. (3) below, the island of {customers} means the basic context about customers, and the island of {steel, concrete, company} shows the basic business context in the mind of people chatting. The bridge"restructuring" shows the company may introduce restructuring, where employees may be fired, for acquiring the good feeling of customers. "Restructuring" might be rare in the communication of the company staffs, but this expresses their potential concern about restructuring in the near future.

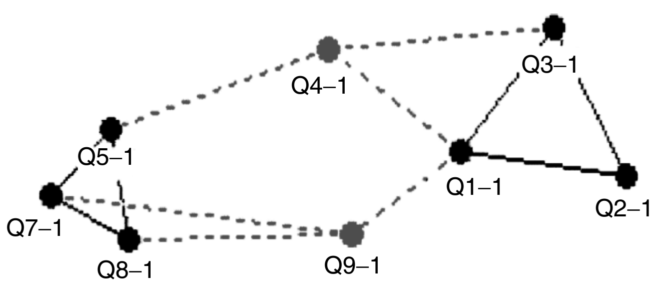

Fig. 3. An example of *KeyGraph on Polaris*: Islands are obtained from $D1$ in Eq. (2), including event-set {Q1-1, Q2-1, Q3-1} and {Q5-1, Q7-1, Q8-1} respectively. The nodes in and outside of islands show frequent and rare events respectively, and Q4-1 and Q9-1 here show rare events in bridges between the two islands.

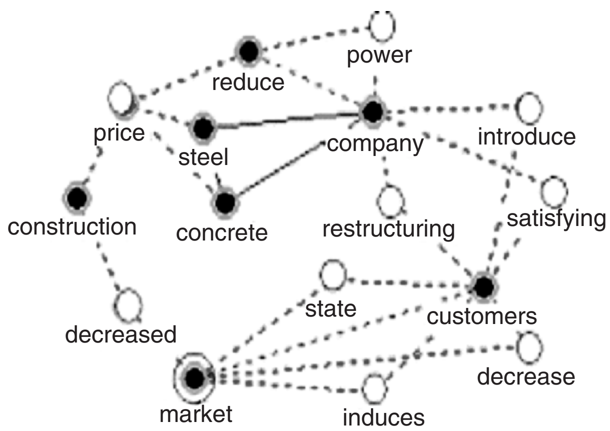

Fig. 4. An example of *KeyGraph*: Islands are obtained from $D2$ in Eq. (3), each including event-set {market}, {steel, concrete, company}, {customers} etc. The double-circled nodes and white nodes show frequent and rare words respectively, forming hubs of bridges.

$D2 =$ *"Mr. X: In the market of general construction, the customers decreased.*
Mr. Y: Yes ... Our company, building from concrete and steel, is in this bad trend.
Mrs. Z: This state of the market induces a further decrease of customers.
Our company may have to introduce restructuring for satisfying customers.
Mr. W: Then the company can reduce the price of concrete, steel, and construction.
Ms. V: But that may reduce us the power of this company." (3)

As in [22], we can also focus on the most interesting part of the dataset by using Boolean search. For example, if we enter "concrete & (market | customer)," the sentences including "concrete" and either or both of "market" or "customer" are chosen from Eq. (3). Then we can read messages about the market or customers of a concrete-production company. As user likes, the extracted sentences can be visualized by KeyGraph, and the user can see the structure of conversations relevant to the market of concrete products.

Influence Diffusion Model (IDM) [14] (see Chapter 7), Whereas KeyGraph visualizes a map of events, commercial items, words, and messages, shown as a set of islands and bridges, IDM shows the influence flows among those items. IDM has been applied to the intra-company informal discussion board, and the manager noticed potential leaders of opinions in her section, and potential desires of the section members [15]. The integration of KeyGraph and IDM made successful discoveries of hepatits scenarios [23].

Since *KeyGraph* appeared, some researchers of chance discovery have been presenting the importance of understanding chances on the bridges among islands, and studying how the bridges can be visualized, extracted and recognized to make a decision, in the corresponding real world domain. At the same time, their most important domain was the communication of human because the tipping points in human society emerge from the communications of people from multiple domains.

In Fig. 5, let us show an exemplification of IDM applied to oral conversation by 8 people discussing about the information ethics on the Internet. Each node here depicts one message, and a thin link with arrow points from a message to one responding to it. The response to a message is identified by choosing a message succeeding many words from the original message. In this figure, islands are the fat clusters of messages surrounded by dotted circular frames.

Each frame has a meaning as marked by the bold large letters. The colored nodes mean influential messages extracted by the method shown below. These colored nodes are extracted by the computation of the *influence* of each node to the overall community discussing. This example shows the bridging parts of the structure make the appearance of a new context in the conversation.

The number on each arrow means the rate of words, inherited from the arrow-tail, of all words in the arrow-head message carried from the original message C1 of which the influence is being considered. These numbers are multiplied along paths from C1 to all other messages following the arrows, and the products obtained for all these paths are summed up as the *influence* of C1.

If message X is more influential according to this computation than its child (neighbor in the lower stream), X tends to have many children. If X

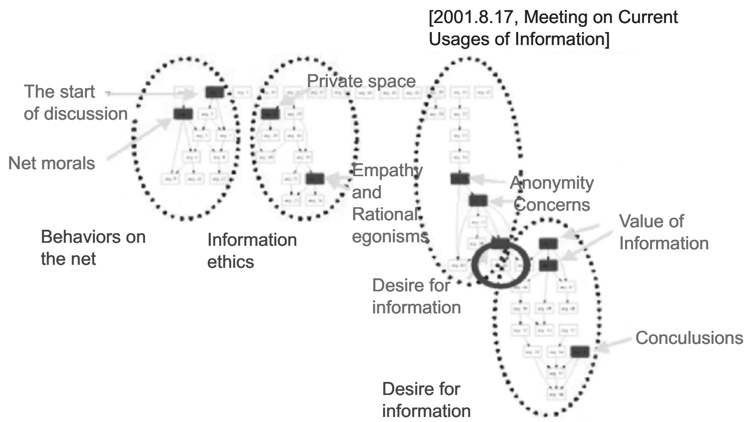

Fig. 5. The islands (dotted circles), bridges (between islands), and the keys (red nodes) in a chain of messages.

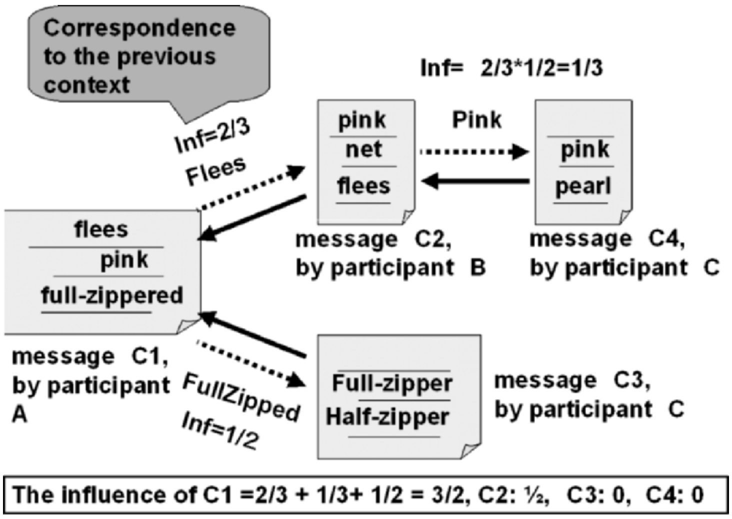

Fig. 6. The computation of influences of messages.

is more influential than its parent, it means the topic of X changed at X and the new topic diffused in the lower stream of X. That is, a message of (locally) largest influences has the strongest impacts on the context-shift in the communication. Such a message is colored as in Fig. 5. By words in such messages, customers can be activated and make expands the market.

For on-line communications in well-designed sites, this method can be used more easily than real conversations because the responding relations among messages can be easily extracted. See Fig. 7. This is the result of IDM for an on-line fan-club of Uniqlo, the biggest brand of casual cloths in Japan. You see the most influential messages are in one path, i.e. the path relevant to pale-olive colored flees. In this manner, the most influential topics tend to appear in a line in a message chain.

In the similar manner as in the chain of messages, we can find the chain of people as in Fig. 8. Here, the word-inheritance between people are numerically dealt with similarly to between messages in Fig. 6, and the people who talked less frequently were sometimes obtained as more influential. As in [3, 4] the results of human chains were validated to be perfectly precise, according to the participant of discussion about architecture.

4.2 Environment for Thoughts and Communications

We can count some previous work on creative communications. In ThinkTank^{TM} [3] (http://pbl.stanford.edu/Research%20Projects/thinktank.htm), messages in a lot of topics are entered in the corresponding thread, and participants can switch between difference threads. For each entry of message, a new thread may be set if user likes to talk in a new context. More

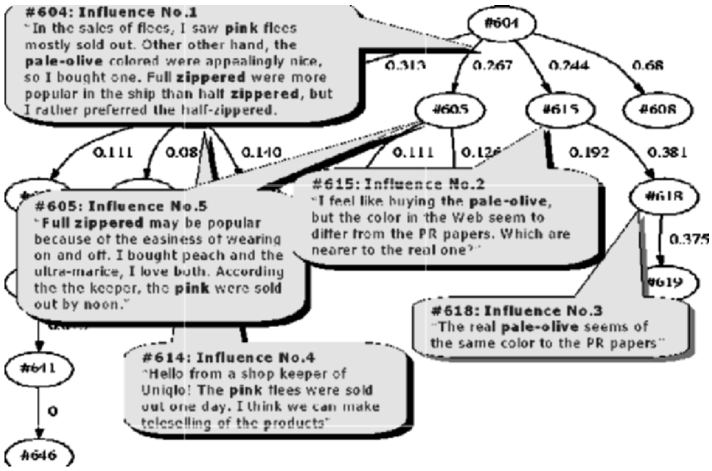

Fig. 7. The chain for an on-line community of fan-club Uniqulo.

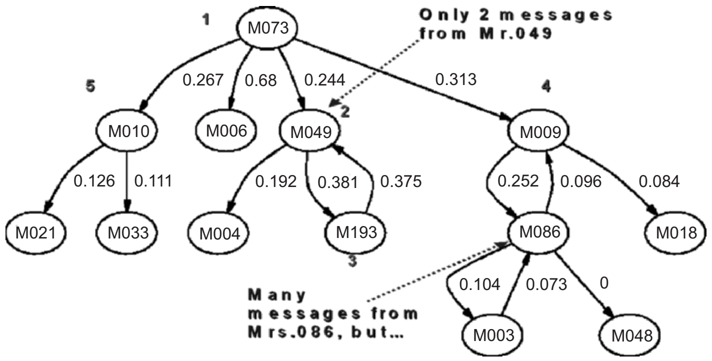

Fig. 8. The human chain from on on-line community.

significantly, data relevant to each message is posted and deposited during the conversations. This way of storing data is apparently useful in reusing the data as external data in Fig. 2, relevant to user's concern, because user's concern is normally expressed in natural words in messages, and such words are linked in ThinkTank^{TM} to data posted by senders. Frucher et al found topics influential to members' future activities from the text of ThinkTank^{TM}, by visualizing with KeyGraph [5].

Fruchter also invented RECALL [2], an integrated tool for aiding design, where various kinds of internal data (data about thoughts and communications) are thrown but are too deep to be saved as explicit knowledge in the mind of designer hem/herself. RECALL is a powerful tool for showing significant pieces of information about some past message he/she feels important but cannot recall explicitly. For example, a group of users keep drawing images about their own ideas, speaking frankly about what they felt and thought, during communication about their business scenarios. These words about intuitive thoughts are hard to write down in a notebook, so they sometimes find it difficult to restore what they discussed, after the session. RECALL allows them to click on a component of the image they drew on the way of communication, and plays the sound RECALL recorded when the users drew the component. They will say "Yes, this is what we thought at that time!" In this book, Fruchter extends her works on social intelligence design (SID), i.e., design of the environment for creative communication, integrating her SID technologies and the concepts on chance discovery.

We also consider the importance of a member's role in a collaborative community. A leader must choose an action that increases benefit and reduces risk. When the leader cannot make such a decision, the group's action will be determined through community member' discussion. However, this decision cannot be made in blind discussions, so that a systematic discussion is necessary to choose effective action in a limited time. Here, the interleaving of divergence and convergence in the discussion according to members' common knowledge and background leads to that effective conclusions. In Part III, Sunayama proposes a bulletin board system framework in which the scenario creation, a series of actions to be undertaken, is established through the discussion and exchange of opinions.

Whereas Llora et al presented the method for reflecting the users' thought onto the output of KeyGraph applied to the text data of communication [12], Iwase and Takama presents a bulletin board system where users can annotate on the map presented by KeyGraph for given external data. They started this work since 2004. In Chapter 17, they show some experiments showing the running of their on-line system. Here, the mapping to the graph generated by KeyGraph and the scenario drawn up by a user is proposed. The mapping is used for extracting the data referred to in the scenario and for annotating those in the original data file. The annotated data files are expected to be used for further data analysis as well as for supporting group discussion.

5 Further Progress with Basic Research

Recently, studies came to be dedicated to experts working in real world domains where discoveries of hidden links are desired. We can say Chance Discovery is one of the leading research domains which began to run from 2000 in this direction.

A relevant research area to Chance Discovery is Evidence Extraction and Link Discovery (EELD), where important links of people with other people and with their own actions are discovered from heterogeneous sources of data [16, 29, 27]. The difference between Chance Discovery and EELD, for the time being, is in the position of human factors in the research approaches. In Chance Discovery, the visualization techniques such as KeyGraph have been used for clarifying the effect of chances, by enforcing the user's thoughts on scenarios in the real environment. On the other hand, the EELD program mainly contributed to identifying the most significant links among items more automatically and precisely than human. I expect these two will meet, because the latest studies in EELD is oriented to coupling symbolic expressions of human knowledge with a machine learning system [28], whereas chance discovery has been integrating the human process of externalizing the tacit experiences and the power of machines for finding a surprising trigger to the activation of the environment. That is, human's interaction with machine intelligence is coming to the centers of these two domains. Some studies in EELD, such as data visualization for decision making [11, 10], serve bridges between human and machine.

However, the complexity of the real world was sometimes even beyond the reach of both Chance Discovery and EELD: A few nerd users of cellular, not frequently sending out comments, are likely to create new fashion causing strong influences on other users. The developer's question is "where is the innovative user?" It is meaningless to ask hundreds of monitors "who gave you the idea to use cellular this way?" because users seldom know the innovative users, but only see other users' accessories of cellular which are the indirect influences of the innovation. As a result, neither comments nor names of innovators can be included in data. Here arose a problem of Data Crystallization.

This may be the meeting point of Chance Discovery and EELD: The detection of unobserved but significant events, as a grand challenge ignited in [24] and presented by Maeno and Ohsawa in Part VI in this book. Data crystallizing means this challenge, and to extend Chance Discovery to the discovery of significant events in more uncertain environment. And, the sphere of real world applications linked from this basic research is expected to include intelligence analysis, development of new products, aiding corporate activities by detecting interest of employees, etc.

For example, let us consider the intelligence analysis, where expert investigators of criminal-group behaviors are exploring missing links among members. The head boss (see the dark guy at the top of Fig. 9) of the criminal organization may phone a few times to sub-leaders managing local sections (Mr. A and Mr. B in Fig. 9). For responding to these top-level commands, each local section holds its internal communication, via different media from that the boss used for contacting sub-leaders. Then, sub-leaders may meet to achieve consensus before responding to the boss. Meanwhile, the boss does not appear in the meetings. In this way, some who is never observed in meetings or mailing lists may be the real boss.

In the study of Data Crystallization, our research team is revealing events potentially important but never observed. For example, some leaders of an online community were deleted from the data of conversation, and our method was applied to the "cleaned" data. As a result, the participants who were most tightly linked to the leaders came to be visualized, and finally the leaders

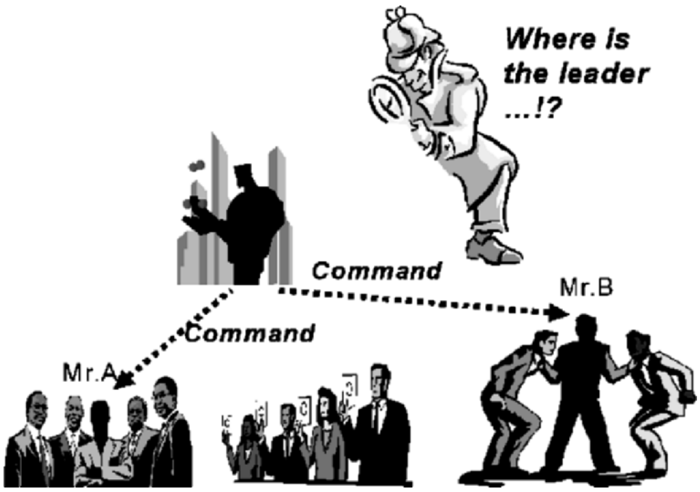

Fig. 9. Intelligence analysis seeking hidden leader.

were identified by investigation of their real human relations. Because such leaders are not included in the data, existing mining methods hardly help in identifying them. Data crystallization is the challenge to this hard problem from an extension of what we have been calling Chance Discovery since 2000.

6 Conclusion of this Chapter

KeyGraph and IDM are not the only techniques for chance discovery. Computation approaches on the Bayesian, the Fuzzy methods that have been developed and employed for dealing with uncertainty in the future are other promising tools as Chai et al and Hubey presents in Chapters in Part II. These may also be mixed with the methods of Data Crystallization.

We should also note that the process of chance discovery is not completed by the Double Helix process. Before everything, the participants of scenario communication should have a prepared mind. In business, one should be prepared for dealing with customers. In all daily life, human has to deal with his/her own and others' value criteria and morals. The Chapters on scenario communication applicable to business by Yada, and on the cognitive aspect introduced by Magnani are two bases for developing a new generation process of chance discovery.

Finally, as pointed out in this Chapter, we should keep in mind that the word "chance" includes the meanings of both opportunity and risk. Whereas the Chapters in Part V show chance discovery methods for finding opportunities for business success, we includes major works on risk discoveries in Part VI. Focusing on the medical topics [Chapter 22 through 24], reader will find how human's attention and concern with risks are to be integrated with the real information in the environment. Chance Discovery, after all, is a research domain about human interaction with the complex and dynamic environment via the exchange of revised information.

References

1. Chance Discovery Consortium http://www.chancediscovery.com (2004)
2. Fruchter, R., Degrees of Engagement in Interactive Workspaces, in Social Intelligence Design International Conference 2003 (2003)
3. Fruchter, R., and Luth, G., CgThinkTank a Web Based Collaboration Tool, Proc. of ASCE Structures Conference, Nashville (2004)
4. Weiss, G.M., and Hirsh, H., Learning to Predict rare Events in Event Sequences, Proc. Int'l Conf. KDD'98, pp.359-363 (1998)
5. Fruchter R., Ohsawa Y., and Matsumura N., Knowledge Reuse through Chance Discovery from an Enterprise Design-Build Project Data Store, The First European Workshop on Chance Discovery (2004)
6. Gaver W.W., Beaver J., and Benford S., 2003, Ambiguity as a Resource for Design, in Proceedings of Computer Human Interactions (2003)
7. Horie, K., Ohsawa, Y., Extracting High Quality Scenario for Consensus On New Specifications of Equipment, 1st Annual Workshop on Rough Sets and Chance Discovery (RSCD) in conjunction with 8th Joint Conference on Information Sciences (JCIS 2005), Salt Lake City (2005)
8. Joshi, M.V., 2002a, Agarwal, R.C., Kumar, V.: Predicting Rare Classes: Comparing Two-Phase Rule Induction to Cost-Sensitive Boosting. PKDD 2002: 237-249 (2002)
9. Joshi, M.V., 2002b, On Evaluating Performance of Classifiers for Rare Classes. ICDM 2002: 641-644 641-644 (2002)
10. Kovalerchuk, B., Relational Text Mining and Visualization, In: Knowledge-Based Intelligent Information Engineering Systems and Allied Technologies, KES 2002. Eds. E. Damiani, R. Howlett, L. Jain, N. Ichalkaranje, IOS Press, Amsterdam, part 2, pp. 1549-1554 (2002)
11. Kovalerchuk, B., Visualization and Decision-Making using Structural Information, In: Proceedings of International Conference on Imaging Science, Systems, and Technology (CISST'2001) 478-484 (2001)
12. Llora, X. et al, Innovation and Creativity support via Chance Discovery, Genetic Algorithms, New Mathematics and Natural Computation, Vol.1, No.4 (World Scientific) in print (2006) (2006)
13. Matsuo Y., and Ohsawa, Y., Finding the Meaning of Clusters, AAAI Fall Symposium on Chance Discovery Technical Report FS-02-01, AAAI Press, pp.7-13 (2002)
14. Matsumura N., Ohsawa Y., and Ishizuka M., Influence Diffusion Model in Text-Based Communiation Poster, The Eleventh Conf. World Wide Web (WWW11), (2002)
15. Kamata, MI., "Software Requirements Elicited through Human-Centric Chance Discovery": Peter MacBury, Yukio Ohsawa and Simon Persons eds., The Discovery and Management of Chance Events, Papers from the AAAI Fall Symposium, AAAI Press (2003)

16. Oates, T., Schmill, M., and Cohen, P.R., Identifying Qualitatively Different Outcomes of Actions: Gaining Autonomy Through Learning . Proc. of the International Conference on Autonomous Agents, pp. 110-111. ACM (2000)
17. Ohsawa, Y., and Fukuda, H., Chance Discovery by Stimulated Group of People, An Application to Understanding Rare Consumption of Food , Journal of Contingencies and Crisis Management Vol.10, No.3, pp.129-138 (2002)
18. Ohsawa Y., and McBurney P. eds, 2003, Chance Discovery, Springer Verlag (2003)
19. Ohsawa Y., 2003a, Modeling the Process of Chance Discovery, Ohsawa, Y. and McBurney eds, Chance Discovery, Springer Verlag pp.2-15 (2003)
20. Ohsawa Y, 2003b, KeyGraph: Visualized Structure Among Event Clusters, in Ohsawa Y and McBurney P. eds, 2003, Chance Discovery, Springer Verlag: 262-275 (2003)
21. Ohsawa, Y., and Nara, Y., Decision Process Modeling across Internet and Real World by Double Helical Model of Chance Discovery, New Generation Computing Vol.21 No.2, pp.109-122, (2003)
22. Ohsawa, Y., Fujie, H., Saiura, A., Okazaki, N., and Matsumura, N., Process to Discovering Iron Decrease as Chance to Use Interferon to Hepatitis B, in Paton, R. (ed) Multidisciplinary Approaches to Theory in Medicine (2004)
23. Ohsawa, Y., 2005a, Scenario Understanding of Hepatitis Progress and Recovery by Annnotation-based Integration of Data based Scenario Maps, GESTS International Trans. Computer Science and Engineering Vol.22, N0.1., pp.65-76 (2005)
24. Ohsawa, Y., 2005b, Data Crystallization: Chance Discovery Extended for Dealing with Unobservable Events, New Mathematics and Natural Computation Vol.1, No.3, pp.373 - 392 (2005)
25. Okazaki, N., and Ohsawa, Y., "Polaris: An Integrated Data Miner for Chance Discovery" In Proceedings of The Third International Workshop on Chance Discovery and Its Management, Crete, Greece (2003)
26. The Danish Board of Technology, 2003, European Participatory Technology Assessment: Participatory Methods in Technology Assessment and Technology Decision-Making,. http://www.tekno.dk/europta
27. Senator, T., EELD Program, http://www.darpa.mil/ito/research/eeld/EELD_BAA.ppt (2001)
28. Sutton, C., Brendan, B., Morrison, C., and Cohen, P.R., Guided Incremental Construction of Belief Networks. 5th International Symposium on Intelligent Data Analysis (2003)
29. Upal M.A., Performance Evaluation Metrics for Link Discovery Systems in Proceedings of the Third International Intelligent System Design & Applications, pages 273-282 (2003)
30. Usui, M., and Ohsawa, Y., Chance Discovery in Textile Market by Group Meeting with Touchable Key Graph, In Proceedings of Social Intelligence Design International Conference (2003)