Hidden Structure Visualization Adaptive to Human's Prior Understanding

Yoshiharu Maeno[1] **Yukio Ohsawa**[2]

[1] Graduate School of Systems Management, University of Tsukuba, maeno.yoshiharu@nifty.com
[2] School of Engineering, University of Tokyo, ohsawa@q.t.u-tokyo.ac.jp

Abstract

This paper addresses hidden structure visualization adaptive to human's prior understanding. The degree of understanding is parameterized by temperature used in the human-interactive annealing process along with stable deterministic crystallization algorithm. This adaptive nature is demonstrated with social network visualization example.

Keywords: Annealing, Chance discovery, Crystallization algorithm, Human-computer interaction, Understanding

1. Introduction

A chance means an event with significant impact on human's decision-making (either opportunity or risk) [1]. To discover a chance means to become aware of and to understand the significance of a chance. It is important especially when the chance is not noticed commonly because of infrequent occurrence [2]. Various social and business domains need understanding of opportunity and risk owing to *chance discovery* [3]. Chance discovery aims at providing a means to *invent the future* rather than to predict the future [4].

However, the researchers of chance discovery have recently recognized a new class of problems where previous tools such as KeyGraph [2] fail to visualize. The problem is to discover important events neither visible nor observed as data. It is learned that there are often invisible events which play an important role in the dynamics of visible events. Such invisible events are named *dark events* after dark matter in cosmology. Promising scenarios shall be put into a concrete shape by understanding the presence, nature, interaction, and the meaning of dark events. This goal has not been within the scope of chance discovery.

Data crystallization [5] is a breaking-through method by Ohsawa, where dummy events which potentially correspond to dark events are visualized. Yet, the complex crystallization algorithm, non-deterministic (depending on the prior knowledge) nature, unstable (sensitive to a small change of input data and algorithmic configuration) posterior event graph, and ambiguity of possible interpretation of the event graph were problems for user's understanding. It has thus been desired that the entire method be separated into two processes as in the chance discovery process; a computer process using a *stable deterministic crystallization algorithm* and human's interpretation process. These processes are re-combined into a process named *human-interactive annealing* [6], [7], [8]. The user can focus the user's particular interest on the combination of user's prior understanding and the quantitative information obtained from the graph the algorithm generated.

In this paper, we address *hidden structure visualization* adaptive to *human's prior understanding*. Visualization can be adjusted based on the degree of the user's prior understanding of the problem domain. The degree is represented by a temperature parameter used in the human-interactive annealing along with stable deterministic crystallization algorithm. The adaptive nature of the annealing process is demonstrated with an example of social network visualization; analysis of persons related to John F. Kennedy's assassination in 1963.

2. Human-interactive annealing

2.1. Annealing

We define a dark event here. A dark event is neither visible nor observable. The occurrence frequency is very small or none. It diffuses randomly like an atmosphere because it does neither tend to cling to a particular event cluster nor tend to appear as a pair with a particular event. In consequence, the co-occurrence is also very small or none. The dark event represents a hidden structure like chain of command from an invisible leader in a terrorist organization.

The human-interactive annealing is similar to the annealing in materials science and simulated annealing. An event graph is used to represent crystallized dark events. The annealing process is a combination of two complementary elements; human's interpretation and crystallization algorithm. The two elements are illustrated in figure 1 with five event graph examples. In the event graphs, the event clusters and dark events are drawn schematically. The dark events are made visible, owing to the crystallization algorithm. The horizontal axis is the number of iteration. The vertical axis corresponds to temperature.

The temperature is a single control parameter representing the degree of human's prior understanding. When the understanding of the problem is believed to be richer, the temperature shall be set higher. More complex higher-order hidden structures shall be revealed. This will lead to the discovery of unique and unexpected scenario. On the other hand, when the understanding is poorer, the temperature shall be set lower. The user should try to understand the basic lower-order structures from the event graph. The iteration in the annealing process is continued until human converges into complete posterior understanding.

As the temperature increases, the following three structural changes occur on the event graph. These are embedded in the annealing process and independent from the stable deterministic crystallization algorithm described next section.

1. Weaker links connecting dummy events are destroyed
2. Weaker links connecting events within a cluster are destroyed
3. The events are divided into larger number of clusters.

The annealing is a combination of human's interpretation and crystallization algorithm. In the crystallization, the computer analyzes the occurrence frequency and the co-occurrence of events. In the heating step, up to the specified peak temperature, the number of clusters and edges between visible events decrease. Weak associations are destroyed. The crystallized dark events disappear. Then, a cooling step comes after the heating step, where event structures are solidified as temperature goes down. The number of crystallized dark events between clusters of visible events increases on an event graph. The clusters are connected to each other to form a single large structure. The crystallization is followed by human's interpretation, where it is checked whether the termination condition is fulfilled or not.

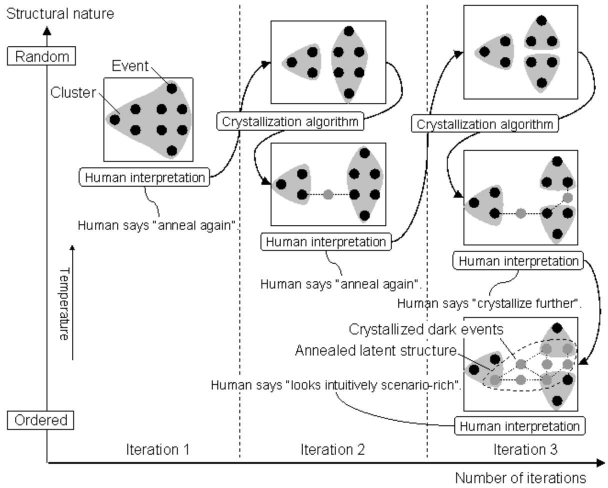

Fig. 1: Human-interactive annealing; iteration of human's interpretation and computer's crystallization of dark events.

2.2. Stable crystallization

A new crystallization algorithm is developed to achieve simple, stable, and deterministic nature. It aims at resolving the complex crystallization algorithm, non-deterministic nature, unstable posterior event graph, and ambiguity of possible interpretation of the event graph. The basic idea of the crystallization algorithm is that visible dummy events are inserted to the basket data to represent dark events artificially [10]. A dummy event is a symbolic expression of a hidden structure containing dark events. Observed basket data $B = \{b_i\}$ shall be the input. The contents of the basket data are a set of events, grouped under a specific subject. In an event graph, a group of dark events can be visualized as a structure representing a bridge between event clusters.

Before processing the basket data, the number of clusters $|C|$ is calculated using the specified temperature, which corresponds to the human's prior understanding. If the algorithm is repeatedly used with the human's interpretation, the temperature is initialized to be a small number and increased. The crystallization algorithm consists of five steps.

The first step is *event identification*. The all events in the basket data $B = \{b_i\}$ $(i \in [0, |B|-1])$ are picked up. A set of the events is denoted by E. An individual event is denoted by e_i $(i \in [0, |E|-1])$.

The second step is *clustering*. The event set E is classified into vertex groups under a specified number of clusters $|C|$. From equation (1), Jaccard coefficient for the all pairs of events is calculated. We employ k-medoid clustering algorithm for computational simplicity [9]. A medoid event $e_{med(j)}$ is an event locating most centrally within a cluster c_j. An event is randomly picked up as a medoid for individual $|C|$ clusters. Other $|E|-|C|$ events are classified into the clusters based on the closeness to the medoids. Then, a new medoid is selected within the individual cluster so that the sum of closeness from events within the cluster to the modoid is maximal. The closeness is evaluated by the equation (2). This is repeated until the medoid converges. The resulting clusters are denoted by $C=\{c_i\}\ (i \in [0,|C|-1])$. This step may employ other clustering algorithms such as k-means or hierarchical clustering [10].

$$\mathrm{Ja}(e_i, e_j) = \frac{\mathrm{Freq}(e_i \cap e_j)}{\mathrm{Freq}(e_i \cup e_j)}. \quad (1)$$

$$\sum_{e_j \neq e_{med(j)},\, e_j \in c_j} \mathrm{Ja}(e_{med(j)}, e_j) \quad (2)$$

The third step is *dummy event insertion*. A dummy event DE_i is inserted into a basket b_i, which results in $b_i \to \{\{e_i\}, DE_i\}$. The index i can be used to identify the basket where the corresponding dummy event was inserted. The dummy event may represent a set of hidden participants to the basket. It may also correspond to the subject to the basket.

The forth step is *co-occurrence calculation* regarding the dummy events. The co-occurrence between a dummy event and clusters is evaluated according to equation (3).

$$\mathrm{Co}(DE_i, C) = \sum_{j=0}^{|C|-1} \max_{e_k \in c_j} \mathrm{Ja}(DE_i, e_k) \quad (3)$$

The fifth step is *topology analysis*. The dummy events are sorted in the order of co-occurrence in step 4. Following this order, the dummy events having co-occurrence with multiple clusters (being a bridge) are picked up. These dummy events are the dark events suggested by the algorithm. The number of links drawn on the event graph calculated using the specified temperature, which corresponds to the human's prior understanding.

3. Human's prior understanding

The adaptive nature of the annealing process is demonstrated with an example of social network visualization; analysis of persons related to a famous historical event; John F. Kennedy's assassination in Dallas, USA in 1963. The input basket data is created, referring to two topics in the encyclopedia Wikipedia; "John F. Kennedy assassination" [11] and "Kennedy assassination theories" [12]. Person's names appearing in a sentence is grouped into a basket. Who is behind the assassination? What is the hidden structure? Some Americans understand the background context very well. Others are familiar with the event. Most non-Americans know only the simple facts. These illustrate the difference in the degree of human's prior understanding.

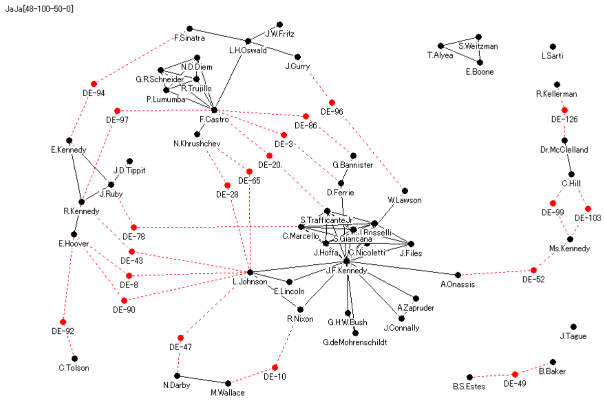

Fig. 2: Visualization of the human network when the temperature is 5 (low).

Fig. 2 shows an event graph visualizing the network when the temperature is 5. Dummy events "DE-x" represents a crystallized potentially hidden person corresponding to dark events. The number of relevant baskets is indicated by "x" in DE-x. When the user do not know about the assassination well, the temperature shall be set low like this example. Small number of large clusters is convenient to get an initial basic understanding. The user should try to understand the basic structure from the event graph. Guessing the identity of crystallized dark events may be relatively easy. There are three big clusters. Hidden persons are indicated between John F. Kennedy and a cluster including communist politicians and Lee Ozwald.

Fig. 3 shows an event graph when the temperature is 14. Fig. 4 shows an event graph when the temperature is 23. As the temperature increases, it becomes more difficult to understand the detailed structured revealed in the event graph. A structure bridging Robert Kennedy, Lyndon Johnson (vice president), Edgar Hoover (FBI director), Nikita Khrushchev, and Fidel Castro is complex, interesting, and mysterious. Are any persons hidden in this area? This question may be difficult to most non-Americans. But it may be stimulating to Americans to get an interesting and unexpected hypothesis in mind.

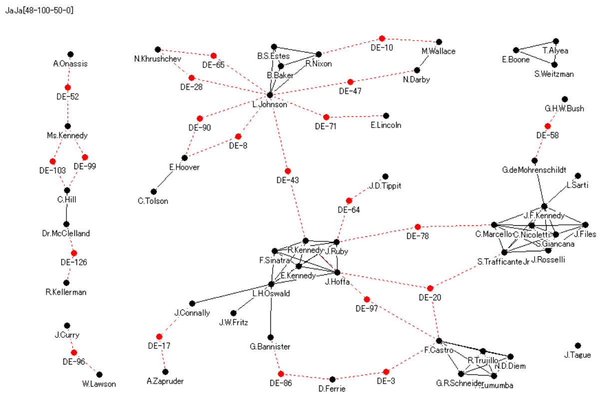

Fig. 3: Visualization of the human network when the temperature is 14 (moderate).

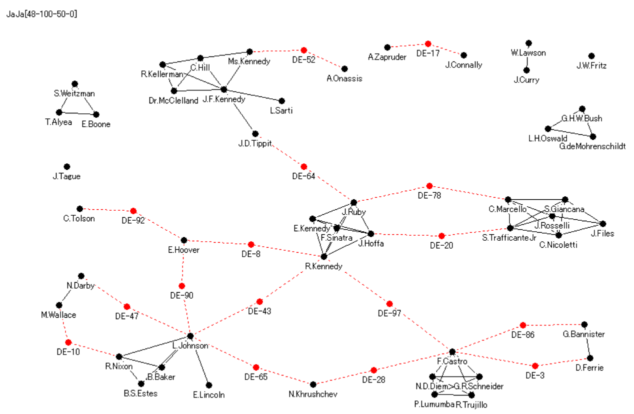

Fig. 4: Visualization of the human network when the temperature is 23 (high).

4. Conclusion

We addressed hidden structure visualization adaptive to human's prior understanding. Visualization can be adjusted based on the degree of the user's prior understanding of the problem domain. The degree is represented by a temperature parameter used in the human-interactive annealing along with stable deterministic crystallization algorithm. The adaptive nature of the annealing process was demonstrated with an example of social network visualization; analysis of persons related to John F. Kennedy's assassination in 1963.

When the understanding of the problem is believed to be richer, the temperature shall be set higher. More complex higher-order hidden structures shall be revealed. This will lead to the discovery of unique and unexpected scenario. On the other hand, when the understanding is poorer, the temperature shall be set lower. The user should try to understand the basic lower-order structures from the event graph. Such adaptive nature is convenient to discover unexpected scenarios in the individual user's own perspective.

5. References

[1] Y. Ohsawa, "Chance discovery for making decisions in complex real world," New generation computing, Vol. 20, pp. 143-163, 2002.

[2] Y. Ohsawa, and P. McBurney eds., "Chance discovery (Advanced information processing)," Springer-Verlag, 2003.

[3] H. Fukuda, and Y. Ohsawa, "Chance discovery by stimulated groups of people: application to understanding consumption of rare food," *Journal of contingencies and crisis management*, vol. 10, pp. 129-138, 2002.

[4] Y. Ohsawa and Y. Nara, "Decision process modeling across Internet and real world by double helical model of chance discovery," *New generation computing*, vol. 21, pp. 109-121, 2003.

[5] Y. Ohsawa, "Data crystallization: chance discovery extended for dealing with unobservable events," *New mathematics and natural science*, vol. 1, pp. 373-392, 2005.

[6] Y. Maeno and Y. Ohsawa, "Crystallization highlighting hidden leaders," Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU), Paris, 2006.

[7] Y. Maeno and Y. Ohsawa, "Understanding of dark events for harnessing risk," Chance discoveries in real world decision making (Springer series in computational intelligence). Springer-Verlag, to be published, 2006.

[8] Y. Maeno and Y. Ohsawa, "Human-computer interactive annealing for discovering invisible dark events," to appear *in IEEE transactions on industrial electronics*, 2006.

[9] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning: Data mining, inference, and prediction (Springer series in statistics)," Springer-Verlag, 2001.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification (2nd edition)," Wiley-Interscience, 2000.

[11] http://en.wikipedia.org/wiki/John_F._Kennedy_assassination

[12] http://en.wikipedia.org/wiki/Kennedy_assassination_theories