2004－ICS－138　(41)
2004／12／7

Concept Map Reflecting Scenario Map
- An Integration of Blood Test Data and Medical Articles -

YUKIO OHSAWA

Abstract: An expert of the domain sometimes needs to catch *scenarios*, where a scenario means a sequence of episodes possible to occur in a certain context, if his/her real-world is very complex and thinking of one episode may not be meaningful. In this paper, a scenario map for the recovery from hepatitis C is visualized with *KeyGraph*, where episodes and bridges between episodes are shown based on a novel causal model. Changes in serum enzymes were linked to interferon, in the scenario. Relevant effects were found for hepatic artery ligation on the concept map from PubMed articles. This shows the enzymes signals the effects of liver treatments.

1. Introduction

According to the definition of "chance" in [2] i.e., an event or a situation significant for decision making, a chance occurs at the cross point of multiple scenarios because a decision is to select one scenario in the future. Here, a scenario is defined as a sequence of events to occur in a certain context. Generally speaking, a set of scenarios forms a basis of decision, in domains where the planning of event-sequence, rather than of a single event, affects the future profits significantly. For example, let us stand on the position of a surgeon looking at a time course or clinical course of symptoms observed in an individual patient. The surgeon should provide the patient with proper treatment at the right time. If he does so, the patient's disease may be cured. Otherwise, the patient's status might be worsened radically. The problem is to choose one from multiple scenarios.

Detecting an event at a crossover point, and selecting the most valuable scenario at such a cross point means a chance discovery. Discovering a chance and taking it into consideration is required for making useful scenarios, but a number of scenarios should be proposed before the chance discovery.

Graduate School of Business Sciences
University of Tsukuba

This motivated the authors to visualize scenario maps, from the real data on blood-test results of hepatitis patients. In this paper, we show the brief abstract of the new findings from scenario maps, and how we reflected the findings to the survey of recent papers on the effects of the changes in enzymes on the recovery of hepatitis.

2. The Data on Blood Tests of Patients of Hepatitis C

Chiba University Hospital has been storing the data of blood-tests for their 771 patients of hepatitis, from year 1982 to 2001. In each test, the values of a suitable part of the 456 given variables were measured. We first cleaned these data in the way below, and obtained 71 variables of which the upper and the lower bound value were predefined, and translated into the form as in Eq. (1) with including interferon.

[The data cleaning in this paper] By the five steps below, we obtained the data of 771 patients, whose blood-test results were recorded for 71 variables.

1) Variables, of which the upper and the lower bounds were given, were chosen from the 456 variables.

2) The values of ions such as K+, Ca2+, Cl-, etc, were deleted because the meaning of these various is ambiguous (they can be the results of drugs not relevant to hepatitis).

3) A value in an unexpected type, e.g., an alphabetical symbol for a numerical variable, and a variable appearing without its value were ignored. Also, if the value

was more than 1000 percent of the upper bound or less than 10% of the lower bound, the value was cut off from the data because the value can be an error, or a state beyond treatment.

4) Values recorded in multiple scales such as [mg/dl] and [mg/l] were put into the same scale, used in the list of the upper and the lower bounds of variables.

5) Some variables were redundant. For example, "jaundice" was identical to T-BIL_high (total bilirubin over its upper bound). Then it was translated into the form of extraordinary value of a numerical variable, i.e., to T-BIL_high.

In Eq.(1), each item represents an extraordinary value or the change in the value of a variable, and one line in Eq. (1) corresponds to one time of blood test. That is, item "X_Y" means an event where the value of variable X took the value or the change as Y shows in 'H' 'L' '+' or '-'. For example, D-BIL_H (L) means a state where the value of D-BIL, i.e., direct bilirubin, was higher (lower) than its predetermined upper (lower) bound of normal value range. And, D-BIL_+ (-) means the value of D-BIL increased (decreased) for two sequent tests. As a result, 284 items came to appear in the data as in Eq.(1). Then, items denoting treatments such as "interferon" were also included in the lines of blood tests (one line in Eq.(1)) during the periods corresponding to the treatments.

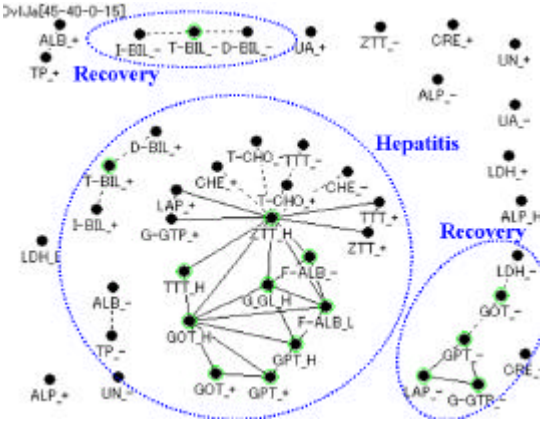

Fig.1. The result of clustering on the co-occurrences between variable-values, for the data of 9314 blood tests of 64 hepatitis C patients who experienced FE_- and FE_L.

D = CHE_- GPT_H GOT_H ZTT_H …,
Interferon ZTT_H GPT_H G-GPT_-...,
G-GPT_- GPT_- CHE_+,... , (1)

It is hard to notice the importance of a low-frequency event in the data, because such an event attracts the attention of a doctor less frequently. Unfortunately, the switch from one episode to another tends to occur in a short time, so its frequency to be detected and recorded in the data comes to be low. Understanding the importance of such a bridging event has been hard as far as we applied previous tools summarized below. For example, the clustering on the overlap value (see next section) between events separates the recovery and hepatitis progress into distinct episodes as in Fig.1.

3. Scenario Map on Three Layered Causality

If a *scenario map*, a map of episodes and the bridges between episodes, is generated on data and presented to the user, the user can draw scenarios on it. For example, a scenario map, where interferon appears in the bridge between the event-clusters corresponding to the two episodes, will help user in understanding the scenario shifting from the episode of progress of hepatitis to the episode of recovery due to the treatment with interferon.

Events in a frequent episode can be shown in a cluster reflecting the co-occurrence, i.e., the tendency of those events to occur in the same line in Eq.(1). Furthermore, it is desired to introduce a new measure for relevance between clusters in a scenario map. In KeyGraph [1,2], as in Figure 2, frequent events are shown in black nodes, and a black link connects a pair of black nodes co-occurring frequently. Here, co-occurrence means two events occur in the same lines in data as Eq. (1). Each cluster, as a result, includes a set of events in a frequent episode. A shift from/to episode, via the occurrence of a rare event, is depicted by a red node and red links bridging the red node and clusters co-occurring strongly with the rare event. See the procedure below.

[The procedure of making a scenario map on KeyGraph]

***Step* 1 (Show episodes as clusters):** The most frequent *M*1 events appearing most frequently (e.g., "GPT_H" in Eq.(1)) are depicted with black nodes, and *M*2 pairs of such frequent events co-occurring the most frequently in the same set (i.e., in the same line in Eq.(1)) get linked with solid black lines. For example, GOT_H, GPT_H, and ZTT_H in Eq.(1) are connected with black lines in Fig.1 and Fig.2. Each connected graph here forms one *cluster*, implying a basis of scenarios, i.e., a set of events in an episode.

***Step* 2 (Show bridges between episodes):** Events co-occurring with multiple clusters, e.g., "interferon" in Eq.(1), 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 or white) than black. We can regard such a new hub as an event which may be rare but significant for episode-switching decisions.

***Step* 3 (Arrows for Scenario Directions):** For showing the time-serial relations between events, thick arrows are attached if the order of increase in the frequency of two states is significant. For example, an arrow is attached from event X to event Y, if Y occurred two sequent times after its absence of two sequent times, after the same pattern of X, for more than 70% of the occurrences of both of X and Y within 10 times of blood test. As a result, the scenario of a typical chrolonological course is expected to appear as a directed path of arrows.

We first proposed *KeyGraph* as a method for extracting keywords from a document by substituting each "event" below with a "word" and each set of events with a "sentence" for Step 1 and Step 2. This worked in extracting a low-frequency word significant for the content structure of the document. Adding Step 3, a scenario map is make from *KeyGraph*.

In computing the co-occurrences at Step 1 and Step 2, we consider three layers of events in the causality model. This model means that T, the cause of the switch from episode A to episode B, is triggered by being required from a previous event S. S can be an event not included in the data, and is supposed to occur from a cause common to events in the preceding situation of A. For example, T means a medical treatment or an immunochemical reaction for overcoming V+, a hazardous external influence, e.g. a virus, causing a set of symptoms in the situation of A. The effect of T is V-, extinguishing the effect of V+ and leading to a stable situation corresponding to B, the recovery.

Between events in the data, the transitional probabilities between events in an episode are caught on the *overlap* co-efficient in Eq. (2). On the other hand, a measure for catching a situational switch can be the mutual(A, T), the *mutual information* given in Eq.(3). These coefficients have been popular in measuring the similarity between words for information retrieval [3].

$$\text{overlap}(U,V)=p(U \text{ and } V)/\min(p(U),p(V)) \quad (2)$$

$$\text{mutual}(U,V)= p(U, V) /p(U)p(V) \quad (3)$$

Here, the mutual information is high if $p(V_+)$ is of low frequency where V_+ is the common cause of A and T. And, the value of mutual(B, T) becomes high if p(T) is low. Depicting the relations between a pair of events of high mutual information, a significant link showing a rare switch, hard to find, is visualized in a red line. Introducing the layers of events reflecting the human perception of events is novel and important in the complex real domain such as progress and recovery of hepatitis.

4. Results of Scenario Maps

Figure 2 shows the result of KeyGraph obtained for the data set of 9314 blood tests. Only the directed arrows corresponding to the bridges (obtained in Step 2) were attached in Step 3 here. The large cluster (bottom of the figure, including GPT_H) corresponds to an event-cluster in the top layer, because they are frequent events connected by large conditional (transition) probabilities. In this cluster, prevalent blood-test results in hepatitis progress are shown. The other episode of recovery is shown in the right side, including GPT_-.

Interferon and other symbols corresponding to treatments were not shown here, but LDH was shown to exist in the switch from the episode of hepatitis progress to the episode of recovery (the peak LDH level signals the beginning of

recovery and the subsequent normalization of the platelet count, which is has been known as a condition for interferon to work [4]). However, the significance of LDH for diagnosis of various diseases came to be considered much later than GOT and GOT [5], because its change does not synchronize with the changes in such established measures as GOP and GPT.

In the result following this, i.e., obtained for the focused data on blood-tests including LDH_low (LDH: lactate dehydrogenase), interferon and the decrease in iron appeared in the bridge between the two clusters (to appear in a forthcoming paper). Here interferon worked between CHE_+ and CHE_-. The recovery (GPT_-, GOT_-, etc) co-occured with CHE_- rather than CHE_+, although CHE has been known to decrease sensitively to the damage to the liver [6,7].

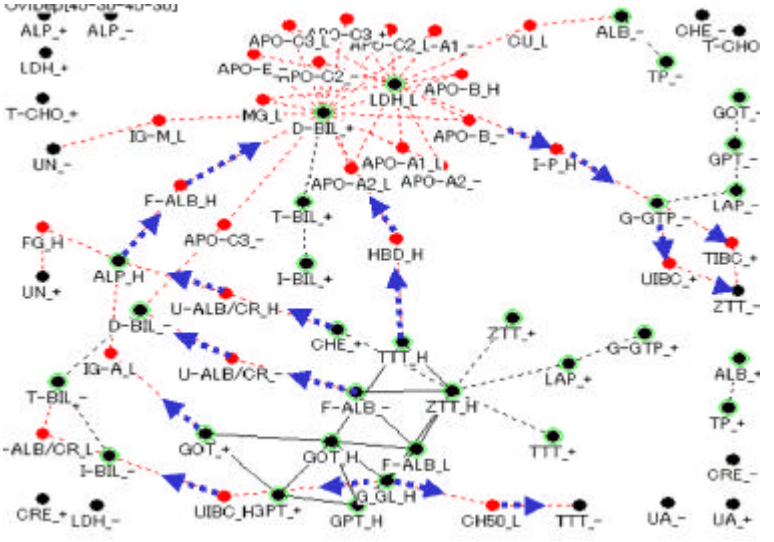

Fig.2 The scenario map for the recovery from hepatitis C, for the same data as dealt in Fig. 1, i.e., 9314 blood tests of hepatitis C. The cluster in the bottom diverges arrows toward the increase in D_BIL and the decrease in LDH_L, which is counter intuitive because bilirubin is expected to decrease with the recovery of hepatitis. The arrows finally lead to the recovery represented by GPT_- and GOT_-.

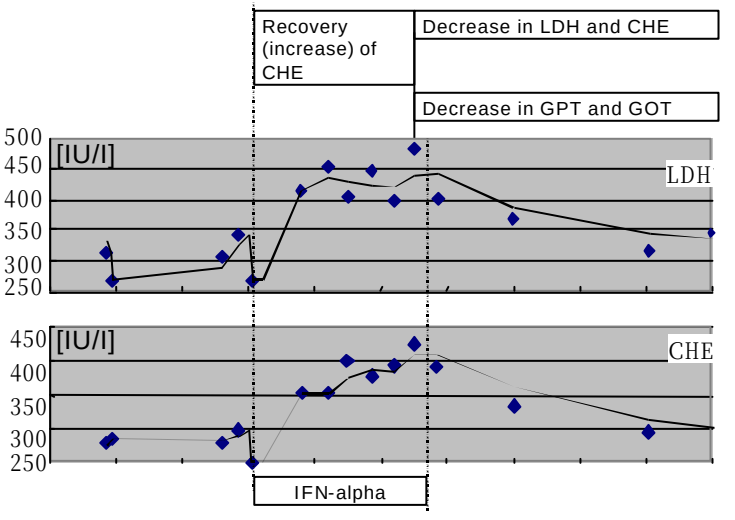

Fig. 3 The correlation between the changes of LDH and of CHE, for one patient.

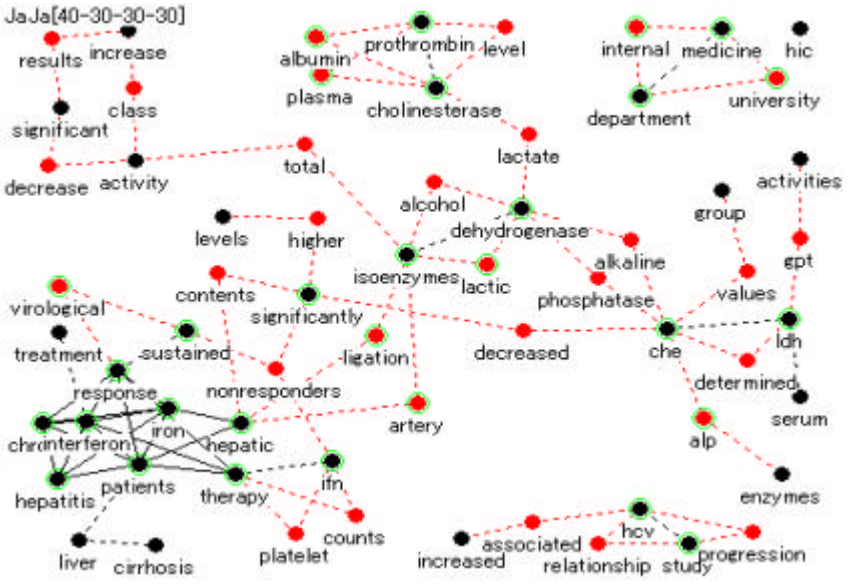

Fig. 4 The KeyGraph result for the collected abstracts of PubMed documents including "LDH" and "CHE."

5. Reflection to Reference Survey on Relevant Enzymes

It is noteworthy that CHE_- replaced LDH_L. The correlation between LDH and CHE as in Figure 2 is surprising because it has been believed LDH increases and CHE decreases in hepatitis. Thus, we retrieved references including "LDH" and ChE" from the abstracts list in the web site of PubMed web (http://www.ncbi.nlm.nih.gov/entrez/).

We collected the abstracts, connected all the data, and visualized it by KeyGraph. Figure 4 shows the result for this. Reading the documents including the terms around words relevant to LDH and CHE in the figure, we find some significant topics relevant to these two terms. For example, isoenzymes such as lactic dehydrogenase (LHD) and choline esterase (CHE), appeared in the center of the figure. Some reports suggested the two tend to differ in a similar manner between in the serum and in the liver [8]. LDH has been reported to increase before decrease in the treatment with interferon [9], or in other treatments as artery ligation [10].

In fact, as in the right of figure 3, choline esterase (CHE) responds to interferon, with the increase and then decrease in its value. For 73% of cases using interferon, CHE increased once quickly after the start of interferon treatment and turned to decrease. This can be relevant to the decrease in CHE as a side effect of surgical operation on the liver, for anti-HCV-positive patients [11].

6. Conclusions

The scenario map on the new causality model visualizes the scenario of recovery, which is a combination of patterns presented in distinct studies. The method with implementing the causality models to *KeyGraph* in making a scenario map is not fully automated, in that the number of visualized nodes and links are given manually and the data is focused on the concern of the user. This freedom for the user is beneficial in arranging the granularity of the output, and helped in understanding switches between frequent episodes, i.e., from hepatitis progress to the recovery. The user was guided to refer to novel reports in PubMed relevant to the finding about new part of extracted scenarios.

References

1) Ohsawa Y., Nels EB, Yachida M, KeyGraph: Automatic Indexing by Co-occurrence Graph based on Building Construction Metaphor, In Proc. Advances in Digital Libraries (1998)

2) Ohsawa, Y., KeyGraph: Visualized Structure

among Event Clusters, Ohsawa, Y. and McBurney eds, Chance Discovery, Springer Verlag pp.262-275 (2003)

3) Boyce BR, Meadow CT, and Kraft DH, Measurement in Information Science, Academic Press (1994)

4) Martin JN Jr, Blake PG, Perry KG Jr, McCaul JF, Hess LW, Martin RW. The natural history of HELLP syndrome: patterns of disease progression and regression. Am J Obstet Gynecol 1991;164(6 pt 1):1500-9.

5) Arlet P et al, Value and diagnostic significance of serum lactic dehydrogenase in internal medicine, : Sem Hop. 1982 Jun 24; 58(25):1543-8 (in French) (1982)

6) Matsuzaki S et al, Abnormalities of serum cholinesterase isozyme in liver cirrhosis and hepatoma (Part II), Gastroenterol Jpn; 15(6):543-9. (1980)

7) Jorge AD and Sanchez D., Serum levels of C3 and cholinesterase in various diseases of the liver., Acta Hepatogastroenterol (Stuttg). 23(1):19-25. (1976)

8) Shaker M et al., Some enzyme interrelationships in serum and organs of Egyptian camel, Arch Exp Veterinarmed. 43(2):241-7 (1989)

9) Nabeshima S et al, Increased lactic dehydrogenase (LDH)-linked immunoglobulin associated with interferon-alpha therapy in a case of chronic hepatitis C, Intern Med. 33(7):446-9 (1994)

10) Nagasue N., Changes in lactic dehydrogenase isoenzymes after hepatic artery ligation in patients with hepatic carcinoma, Gastroenterol Jpn. 1975;10(2):150-6 (1975)

11) Murakami S et al, Changes in liver enzymes after surgery in anti-hepatitis C virus-positive patients, World J Surg. 28(7):671-4. (2004)