Identity Tracking of a Disease as a Causal Chain

Kouji Kozaki[1,*], Riichiro Mizoguchi[1,*], Takeshi Imai[2] and Kazuhiko Ohe[2]

[1] ISIR, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, Japan

[2] Department of Medical Informatics, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan

ABSTRACT

This paper discusses the ontological treatment of identity of a disease based on an ontological model of diseases as causal chains. An individual causal chain changes through its processes such as extending, branching, fading, etc. When we suppose that an individual disease is constituted by such a causal chain, changes of the causal chain imply changes of the disease. Then, how should we deal with identity of a disease? To answer this, we consider identity of a disease as a causal chain based on some ontological theories. As a result, we propose an ontological model for identity tracking of a disease as a causal chain. This work should contribute to the ontological treatment of identity of diseases in medical information systems.

1 INTRODUCTION

Recently, many medical ontologies have been developed for realizing sophisticated medical information systems, such as OGMS (Scheuermann et al., 2009), DOID (Osborne et al., 2009), and IDO (Cowell et al., 2010). They mainly focus on the ontological definition of disease with related properties. We proposed a definition of a disease involving capturing a disease as a causal chain of clinical disorders (Mizoguchi et al., 2011). This paper discusses the evolution of individual diseases, whereas the main concern in our previous work was capturing static characteristics of diseases. An individual disease undergoes changes as it evolves, and some of these changes may cause new diseases to appear as a result. In some cases, some symptoms may appear as aftereffects lingering beyond the point when the original disease was cured. Even if the disease can be cured completely, it could appear in the patient's anamnesis. Here, the problem is how these changes should be dealt with based on our disease ontology. We discuss it from the viewpoint of identity tracking of a disease based on an ontological model of causal chains and a theory of identity (Kozaki et al., 2010).

This paper is organized as follows. The next section outlines the definition of a disease in our previous work. In Section 3, we describe an ontological theory of identity tracking of a disease. In Section 4, we summarize some related work. Finally, we present concluding remarks with a discussion of future work.

2 A MODEL OF DISEASES

After it begins to exist, a typical disease, as a dependent continuant, enacts extending, branching, and fading processes before it disappears. Thanks to these processes, a disease can be identified as a continuant that is an enactor of those processes. Such an entity (a disease) can change according to its phase while keeping its identity. On the basis of this observation, we defined a disease as:

Definition 1: Disease (Mizoguchi et al., 2011)
A disease is a dependent continuant constituted of one or more causal chains of clinical disorders appearing in a human body and initiated by at least one disorder.

Note that, although any disease has dynamic flows of the propagation of causality as its internal processes, it is the enactor of its external processes, such as branching and extending its causal chain of disorders.

When we collect individual causal chains belonging to a particular disease type (class), we are able to find a common causal chain (partial chain) that appears in all of the instance chains. By generalizing such a partial chain, we obtain the notion of a core causal chain of a disease as follows:

Definition 2: Core causal chain of a disease[1]
A sub-chain of the causal chain of a disease whose instances are included in all the individual chains of all instances of a particular disease type. It corresponds to the essential property of a disease type.

Definition 2 provides a necessary and sufficient condition for determining the disease type to which a given causal chain of clinical disorders belongs. That is, when an individual causal chain of clinical disorders includes instances of the core causal chain of a particular disease type, it belongs to the disease type. We can thus define such a disease type that includes all possible variations of physical chains of clinical disorders observed for patients who contract the disease. According to a standard definition of subsumption, we can introduce an *is-a* relation between diseases using the chain-inclusion relationship between causal chains.

Definition 3: *Is-a* relation between diseases
Disease A is a supertype of disease B if the core causal chain of disease A is included in that of disease B. The inclusion of nodes (clinical disorders) is judged by taking an *is-a* relation between the nodes into account, as well as sameness of the nodes. (Mizoguchi et al., 2011)

Assume, for example, that (non-latent) diabetes and type-I diabetes are respectively defined as *<deficiency of insulin → elevated level of glucose in the blood>* and *<de-*

---

* To whom correspondence should be addressed: {kozaki, miz}@ei.sanken.osaka-u.ac.jp

[1] This is a revised version of Definition 3 in (Mizoguchi et al., 2011).

struction of pancreatic beta cells → lack of insulin I in the blood → deficiency of insulin → elevated level of glucose in the blood>. Then, we get *<type-I diabetes is-a (non-latent) diabetes>* according to Definition 3.

There is another way to define *is-a* relations among disease types. Assume that the core causal chain of arteriopathic disease is *<disorder of artery → disorder of heart function>*. Then we can define a subtype disease of arteriopathic disease which has the core chain *<disorder of aorta → disorder of heart function>* by specializing disorder of artery to disorder of aorta. The specialization of disease types in the latter way is always sound because the new specialized types include only instances of the original types via chain-inclusion (the former way). On the other hand, generalization in the latter way does not always generate sound results via chain-inclusion because what we get will include causal chains that do not go through the core causal chain of its supertype.

We understand there is a difficulty in defining *main pathological/etiological condition(s)* for each disease due to the variety of disease manifestations. However, at least we should be able to expect the existence of something common to all instances of a disease type, otherwise we cannot define the type. The core causal chain is such a common thing that contains something corresponding to the *main pathological/etiological condition(s)*. That is, we would like to claim that the right terminal end (the disorder located at the most downstream position) of the core causal chain of a disease defined in a bottom-up way from a collection of individual causal chains should be understood as a kind of *main pathological/etiological condition(s)*. In an extreme case, the core causal chain of the root disease of an *is-a* hierarchy would consist of a single clinical disorder as a special case of a causal chain, and this should correspond to the *main pathological/etiological condition(s)* of the disease.

3 IDENTITY TRACKING OF A DISEASE

A disease undergoes change during its course. The cause may disappear, especially in its latter phase. Such disappearance might have a negative influence on determination of the disease of individual causal chains. Many other fluctuations of disorders might influence the causal chains, too. Ontology of diseases should address such issues. One of the major issues related to such observations is identity tracking of diseases. To tackle this identity problem, we need to formulate the identity of a causal chain of clinical disorders. Because the disease is a continuant, it can change while maintaining its identity. The theory of processes and events adopted in this paper is found in the literature (Galton & Mizoguchi 2009), in which the two key notions are (1) processes are intrinsically ongoing/in-progress whereas events are not, and (2) an event is constituted of processes.

3.1 Classification of causal links

There are two kinds of processes (Mizoguchi et al., 2011):

(1) Cumulative continuous process: a process that proceeds without completing the current process at every instant in time.
(2) Non-cumulative process: a process that proceeds by completing the current process at every instant in time.

A causal chain is composed of one or more pairs of entities such as a causal event/process and an effect event/process, where the latter has been caused by the former. The effect becomes another cause that causes another effect in the case of multi-pair chains. We define three kinds of causal links from the perspective of "ongoingness".

Definition 4: Ongoing causal link
This is a link that shows that the propagation of causality from the cause to the effect is ongoing.

Definition 5: Historical causal link
This is a link that shows that the propagation of causality from the cause to the effect has terminated.

Definition 6: Pseudo-simultaneous causal link
This is a link that shows only a pseudo-simultaneous relation between cause and effect, excluding the temporal information. It is used mainly for causal events that can be viewed as having happened at the same time.

Continuous propagation of causal effect requires that both causal and effect processes are ongoing. Therefore, ongoing links can appear only between ongoing processes, and they correspond to causal relations found in cumulative continuous processes; *<braking → decelerating>* is a good example of this type of causal link. We believe that many of the causal links found in diseases are also of this type. For example, in the chain *<deficiency of insulin → elevated level of glucose in the blood>*, both processes are definitely ongoing. The same applies to *<angiostenosis → lack of oxygen in myocardial cells → necrosis of myocardial cells>* which is a causal chain in myocardial infarction. On the other hand, *<loss of kidney function → increase of waste in the blood>* might be problematic, since "loss of kidney function" is a past event and hence not ongoing. It is true that *"loss of kidney function"* is an event, but after the event, a new process or state *"kidney is not functioning"* is ongoing, and this causes *"increase of waste in the blood"*. In reality, a historical causal link appears between *"loss of kidney function"* and *"kidney is not functioning"*, and an ongoing link appears between *"kidney is not functioning"* and *"increase of waste in the blood"*. The same applies to *"virus invasion", "steroid treatment"*, etc.

Readers might suspect that we have to write unnecessary processes as follows: "finish traveling" and then "not traveling". Of course, this is not the case. We are describing causal chains of clinical disorders which we believe explain a disease under consideration, and hence we can select what to write as a disorder. Thus, we do not have to write processes that are not necessary for explaining diseases.

Although many of the causal links appearing in diseases are ongoing, historical causal links also should appear in

such cases where the causal events have terminated and the causality does not flow currently. A typical example is the event of a virus invasion that had caused an infection. Another case is a cured disease consisting of a causal chain of past events, which will be discussed in Section 3.2 in detail.

Transition from an ongoing process to a terminated event can be explained in terms of the change from ongoing causal link to historical causal link as follows: when the causing process has terminated, the process constitutes a new event as a past event of the causation, together with a change of the ongoing causal link to a historical causal link.

A pseudo-simultaneous causal link is special. Let us assume a series of events associated with a vase, such as "fall onto the floor, collision with the floor, break, and spill water". It is a common phenomenon and corresponds to a non-cumulative process. We take two events from it: collision and breakage events. It seems like the two events happened at the same instant in time. Precisely speaking, however, these two events need non-zero time intervals, and the collision happened before the breakage. In spite of this reality, people tend to see that they happened at the same time with the understanding that the collision caused the breakage. A pseudo-simultaneous causal link is introduced to represent such a causal relation. It does not include a non-zero temporal interval but deals with the two events as if they happen at the same time while keeping the causal order. It also can be interpreted as a degenerated causal link that includes the change of an ongoing causal link to a historical causal link and is useful for representing causal events composed of non-cumulative processes.

3.2 Classification of causal chains

We introduce three kinds of causal chains to capture the change of diseases according to the causal links as follows:

Definition 7: Strongly connected causal chain
Any causal chain all of whose causal links are ongoing causal links; that is, it is composed of only ongoing processes. A causal chain composed of one ongoing process is included as a special case.

Definition 8: Weakly connected causal chain:
Any causal chain that includes at least one historical or pseudo-simultaneous causal link and at least one ongoing process.

Definition 9: Historical causal chain:
Any causal chain all of whose causal links are historical or pseudo-simultaneous causal links. That is, it is a causal chain composed only of terminated events.

Definition 10: Ongoing causal chain:
Any causal chain that has at least one ongoing process. It subsumes the strongly connected and weakly connected causal chains.

When an ongoing process of a strongly connected causal chain has terminated, the whole chain becomes a weakly connected causal chain composed of two strongly connect causal sub-chains. If all of the ongoing processes have ter-minated, then the chain becomes a historical causal chain. Note here that whatever change happens, the identity of a causal chain remains the same as before it becomes a historical causal chain. This topic will be discussed in Section 3.3.

3.3 Identity of a causal chain

We begin our discussion on identity of a causal chain by talking about identity of entities (Kozaki et al., 2010). There are two types of identity:

Instance identity: Identity for discussing the sameness of instances.
Class identity of a thing: Identity for discussing the sameness of the class to which the entity belongs.

In order to talk about identity of a disease instance, we need to establish identity of a causal chain of clinical disorders. So, what we should discuss is under what conditions the instance of a causal chain loses identity.

After it has been born, an individual causal chain of clinical disorders grows, branches, etc. and it may reach another disease. Thanks to appropriate remedies, its causal disorders disappear, whereas some downstream disorders are still ongoing. During such a change, the disease, that is, the causal chain of clinical disorders, keeps its identity. For example, even if diabetes in a patient causes a cerebral infarction, the clinical causal chain remains the same; that is, no new causal chain appears because of it. Furthermore, even if an after-effect remains after all of the main causes have disappeared, we should be able to identify it as the same causal chain in a different state. In order to deal with such changes properly, we employ Identity for replacement (*I-Rep*) among the four kinds of identity discussed in our previous work (Kozaki et al., 2010). *I-Rep* is defined as:

***Identity for replacement* (denoted as *I-Rep*):**
Identity wherein an instance of the whole continues to be itself without becoming another thing while its parts are being replaced, independently of the kinds and number of replaced parts.

A typical example for the use of *I-Rep* is a bike whose parts are replaced one by one. Whatever parts are replaced and in whatever number, a bike whose parts are replaced remains the same in terms of *I-Rep*. Exactly the same applies to causal chains of clinical disorders. Some readers might notice that we could use ***Identity for essentiality*** (*I-Ess* for short) instead of *I-Rep*. *I-Ess* is defined as identity determined by the essential property of an individual. If you use *I-Ess* with, say, a saddle that you love as its essential property and parts of your bike are being replaced as in the case above, then you would say, "This is not my bike any more," when the saddle has been replaced with a new one. In the case of diseases, for example, an individual causal chain of diabetes has the essential property of an elevated level of glucose in the blood. If its identity is defined in terms of *I-Ess* for that property, then it would become another causal chain when the glucose level becomes normal

because it loses its identity. However, this is not appropriate because identity of the causal chain itself should be kept unchanged as long as some disorders caused by the diabetes remain. Furthermore, even if only an aftereffect is left, it should be identified as the same causal chain in a different state. This is why we employ *I-Rep* instead of *I-Ess*.

Class identity, on the other hand, is used to talk about to what type an individual causal chain belongs. As stated in Definitions 1 and 2, an individual causal chain of clinical disorders *Ci* is identified as belonging to a disease type D_A when it includes an instance of the core causal chain of D_A as a part, and then, it gains class identity for D_A. This *Ci* is said to be an ongoing disease D_A if it is an ongoing causal chain. *Ci* can get another identity for a different disease when it also includes an individual core causal chain of another disease, say, D_B. Class identity is kept until the ongoing causal chain becomes a historical causal chain. Therefore, the extinction condition of causal chains in terms of *I-Rep* and *I-Ess* is the same.

Let us discuss conditions for extinction of a causal chain in terms of *I-Rep*. Any individual causal chain, *Ci*, is born as a single ongoing process, that is, as a strongly connected causal chain. It never changes its *I-Rep* identity by replacement of its parts. Neither growth nor branching can change its identity. When an intermediate disorder terminates to become an event, it becomes a weakly connected causal chain, keeping its *I-Rep* identity. In other words, it does not change its *I-Rep* identity if it has at least one ongoing disorder. Now, we discuss what happens when all the ongoing processes have terminated and hence all the links have become historical causal links. According to previous work (Galton and Mizoguchi, 2009), an event is constituted of processes, and when all associated processes have terminated, it finally appears as a whole as the result of constitution of past (terminated) processes. Because processes and events are different categories, the newly created event should have different identity from that of the processes. When all the links have become historical causal links, therefore, the resulting historical causal chain must be a different instance (chain) from the ongoing causal chain because all of its nodes and links are different from the original chain (chain before the change). That is, when all disorders included in an ongoing causal chain have gone (been successfully treated), the chain loses its identity and becomes a historical causal chain of different identity, which corresponds to curing the disease, and the resulting historical causal chain represents the anamnesis of the patient. Those past processes constituting the events represent how the previous disease appeared in the patient's anamnesis and how the disease was cured.

Note here that a change of part of the ongoing causal chain does not influence the identity of the whole chain thanks to the *I-Rep* identity; that is, an ongoing causal chain remains the same chain after curing a few disorders in the chain before all of the disorders have been cured.

We associate two kinds of identity with each individual causal chain of disorders, namely, instance identity and class identity. At first glance, class identity seems to be useful to talk about when a disease is cured. However, this is not the case, because a causal chain keeps the same instance identity and remains being the same disease as long as there are still ongoing downstream disorders/symptoms even when all the disorders included in its core causal chain have been cured. When all disorders included in a causal chain are cured, the ongoing causal chain disappears and a new historical causal chain is born. As discussed already, because the newly born historical causal chain is a different individual from the corresponding ongoing causal chain, it should have a different instance identity. However, its class identity should be the same as that of the original causal chain, since a previous disease that appeared in the patient's anamnesis must be the same as the disease of the corresponding ongoing one. In summary, the historical causal chain should have the same class identity as the original disease. Considering various cases concerning ongoing/historical states, there can be several states for a causal chain. We can identify the following three major states:

1) Ongoing: an individual causal chain *Ci* of disease D_A whose core causal chain is still ongoing.
2) Curing: *Ci* as a whole is ongoing but the core causal chain of disease D_A of *Ci* is inactive (has become a historical causal chain).
   2.1 A few partial causal chains in the downstream are still ongoing.
   2.2 No ongoing causal chains but a few long-term remaining disorders that have no influence on others (aftereffects).
   2.3 Although disease D_A seems to have been cured at a certain time, it might recur in the future.
   2.4 In a case of cancer, it might metastasize to other regions, etc.
3) Cured: *Ci* as a whole is a historical causal chain.

The first and last cases are obvious. The second case is vague in its nature. It may not be clearly defined, especially when doctors consider aftereffects, recurrence, metastasis etc. We would like to cover such cases as states in which the core causal chain becomes a historical causal chain while the chain as a whole is ongoing. However, there is room to investigate more-detailed classification of these cases.

3.4 Identity tracking of a disease

As a summary of the discussion made thus far, we discuss examples of the life of an individual disease as a causal chain of clinical disorders, as shown in Fig. 1. Note here that we are not trying to predict the future state of a particular disease of a particular patient, but to exemplify possible changes of diseases in the real world represented as a causal chain of clinical disorders in terms of its *I-Rep* identity and type membership.

We do not discuss cases where pseudo-simultaneous causal links appear, since these should be dealt with in a similar way to the case of historical causal links.

(1) A causal chain of clinical disorders, Ci, is born as a strongly connected causal chain. At the same time, it gains its instance identity ($I_{id\_}1$). Ci can extend, branch, etc. while keeping its identity. Before it establishes itself as an instance of a particular disease type, its class identity remains undetermined.

(2) When Ci extends and includes an instance of the core causal chain of the disease *diabetes* as its partial chain, it becomes an instance of *diabetes* and gets class identity $C_{id\_}A$ for *diabetes*.

(3) Once Ci gets class identity for a disease type, it never loses it until Ci disappears, and the class identity remains unchanged independently of its extension and/or branching because it still includes the instance of the core causal chain of the disease.

(4) Ci can get more than one class identity, say, $C_{ins\_}B$ for *diabetic retinopathy*. Ci may extend further and cause serious disorders/symptoms, which could lead to the patient's death.

(5) After some changes of the disorders as well as the termination of the ongoing process to become an event, related causal links of Ci change from ongoing links to historical links. Then, Ci changes from a strongly connected causal chain to a weakly connected causal chain. Nevertheless, Ci keeps its instance identity as an ongoing causal chain.

(6) When a partial chain of Ci corresponding to the core causal chain of *diabetes* and *diabetic retinopathy* becomes a historical causal chain (the associated disorders have disappeared), we might say that the patient is cured of the disease. However, in some cases, a few disorders/symptoms might still remain, and hence Ci might still be ongoing. For example, some destroyed nerve tissue might be left after that change.

(7) When all of the processes in Ci have terminated and become events, and hence all of the causal links become historical links, Ci becomes a historical causal chain as a new individual causal chain, all of whose nodes are events constituted of related past processes.

4 CONCLUSION

We have discussed and proposed an ontological account of the evolution of a disease as a causal chain of clinical disorders. As far as the authors know, there is little activity regarding ontology for the evolution of diseases in DOID and OGMS, whose main focuses are definitions of related types. Although OGMS defines "disease course", it is dealt with as a process, whereas we discussed it in terms of a causal chain as a dependent continuant. We believe that, together with our previous paper (Mizoguchi et al., 2011), we have made a step toward a comprehensive ontological account of diseases, covering their evolution. Other researchers (Galton et al., 2005) have discussed causal relations amongst events, states, and process. Although our notions about the kinds of causal links share some ideas with theirs, the definition of identity of a causal chain is a unique feature of our theory.

ACKNOWLEDGEMENTS

REFERENCES

Scheuermann, R. H., Ceusters, W., and Smith, B. (2009) *Toward an Ontological Treatment of Disease and Diagnosis.* Proc. of the 2009 AMIA Summit on Translational Bioinformatics, 116-120, San Francisco, CA.

Galton, A. et al. (2005) *Processes and Events in Dynamic Geo-Networks.* Proc. of GeoS 2005, LNCS 3799, 45–59, Mexico City, Mexico.

Galton, A. and Mizoguchi, R. (2009) *The water falls but the waterfall does not fall: New perspectives on objects, processes and events.* Applied Ontology **4**(2), 71-107.

Osborne, J. D., et al. (2009) *Annotating the human genome with Disease Ontology.* BMC Genomics 10(1):S6.

Cowell, L. G. and Smith, B (2010) *Infectious Disease Ontology.* Infectious Disease Informatics, Chapter 19, Sintchenko V., 373-395.

Mizoguchi, R., et al. (2011) *River Flow Model of Diseases*, Proc. of ICBO2011, 63-70, Buffalo, USA.

Kozaki, K., Endo, S., and Mizoguchi, R. (2010) *Practical Considerations on Identity for Instance Management in Ontological Investigation*, Proc. of EKAW2010, LNAI6317,16-30, Lisbon, Portugal.
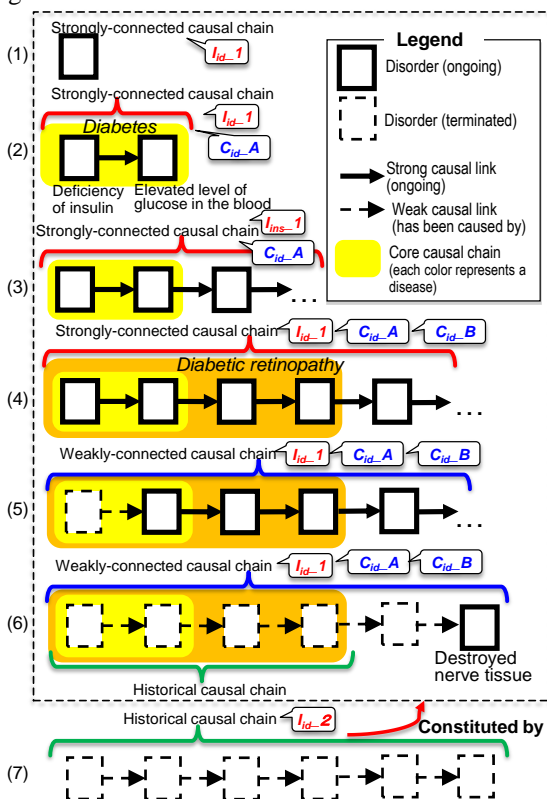

Fig. 1. An example of the life of an individual disease as a causal chain of clinical disorders.