

Knowledge Infrastructure for Knowledge Sharing among Patients, Doctors and Researchers

Tetsuya Maeshiro and Shin-ichi Nakayama

University of Tsukuba
1-2 Kasuga, Tsukuba
Ibaraki, 305-8550 Japan

Yuri Ozawa

Saitama Social Insurance Hospital
4-9-3 Kitaurawa, Urawa
Saitama, 330-0074 Japan

Abstract

We are conducting a project to build a knowledge infrastructure to improve common understandings and knowledge among doctors, patients and researchers. The knowledge infrastructure consists of terms and semantic relationships among them, represented using the hypernetwork model. In order to build a merged knowledge representation, the terms used by the patients and doctors/researchers were analyzed. Less than fifth of terms were common, indicating differences in viewpoints.

We are conducting a project to build a knowledge infrastructure to improve common understandings and knowledge among doctors, patients and researchers. The primary objective is to improve the medical treatment quality by enhancing doctor-patient relationship from the mental aspect. Currently, the target disease is lifestyle-related illness, specifically diabetes.

The knowledge infrastructure connects three types of people through the knowledge of the target disease. They are (i) patients and their families, (ii) doctors and (iii) Medicine and Biology researchers. Although they share and explore the same knowledge, the required facets are different, and their objectives are different as well. Some knowledge is required only by one of three person types, and others are accessed by all three types (patients, doctors and researchers). The information flow direction should also be considered. Existing flow directions are: (1) Patients \Rightarrow Doctors and Researchers; (2) Doctors and Researchers \Rightarrow Patients; (3) Researchers \Rightarrow Doctors; and (4) Doctors \Rightarrow Researchers. These facts should be considered when representing knowledge, and adaptive representation mechanism is necessary.

With the obligation of informed consent and the difficulty on understanding by patients due to accelerated advance of medical techniques, many research have focused on explanation methodologies, texts and visual aids for patients to understand better the explanations provided by their doctors. However, these aspects are not of fundamental importance to improve the doctor-patient relationship. Our experiences indicate that solely improving the explanation quality of doctors to patients does not solve the communication problem

between doctors and patients. Empirically, this communication problem is inversely proportional to the strength of trusting relationship. Without trust, no communication is established.

Our project focuses on mechanisms and methodologies to build knowledge infrastructure that improve mutual understandings and knowledge sharing, and properties that the infrastructure should possess.

The knowledge is represented using the hypernetwork model(?), extended from hypergraph (Berge 1989), which has more representation power than conventional knowledge representation models that are based on graph (Berge 2001). The main difference is the capability to represent N -ary relationships and the property of duality. The proposed model follows basic definitions of semantic networks (Quillian 1967), where a node is connected to other nodes (1) to specify the nodes or (2) when nodes are related by some relationship.

A graph G is defined as $G = V \times V$, where V is the set of nodes, and edge connects a pair of nodes. As the formula indicates, a graph is a matrix generated by product of nodes. On the other hand, the hypergraph H is defined as $H = V \times E$, where E is the set of edges. $V \times E$ is a matrix that results from product of vertices by nodes. Hence an edge of hypergraph, called hyperedge or hyperlink, connects any number of nodes. Edges of hypergraph connecting more than two nodes are represented with circle surrounding the connected nodes. This is a notable difference from graph that can connect only a pair of nodes. Note that N -ary relationship is fundamentally different from the combination of binary relationships, as they represent distinct concepts. Another interpretation of the formula $V \times E$ is under the framework of set theory, where each hyperlink defines a distinct subset of nodes.

Another representation of hypergraph, the Bipartite graph, is obtained by converting the formula $V \times E$ to a formula similar to the graph, $V \times V$. If the set of links E of hypergraph is treated as a set of nodes V_E , $H = V \times E$ becomes $H_B = V \times V_E$, a structure called bipartite graph. Two different types of nodes exist in bipartite graph. One is the set of nodes that exists as nodes (vertices) in original hypergraph, denoted vertex nodes. The other one is the set of hyperlinks in original hypergraph, denoted link nodes. Although bipartite graph has expression capability equivalent

	Vertex	Relation	Attribute
Vertex	—	Connect	Connect
Relation	Connect	—	Connect
Attribute	Connect	Connect	—

Table 1: Connectivity among vertex node, relation node and attribute node. Vertex denotes Vertex node, Relation denotes Relation node, Attribute denotes Attribute node.

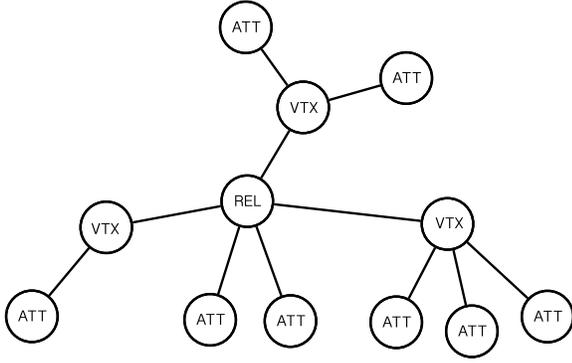


Figure 1: Connectivity of hypernetwork model. VTX denotes a vertex node, REL a relationship node, and ATT an attribute node.

to the hypergraph, it is a graph because a link always connects a pair of nodes. However, a link never connects nodes of the same type, i.e., two vertex nodes or two link nodes. A connection is allowed only between a vertex node and a link node.

The third type of node exists in the proposed model, denoted attribute node, which has no correspondence in hypergraph and subsequently in bipartite graph. Due to the presence of this node, though a hypergraph can be converted to a bipartite network, this process is irreversible.

In the context of knowledge representation models, a concept or an entity represented by vertex node can be specified in two ways: by describing the attributes of the vertex node, or by relating to other vertex nodes. Combination of the two is also possible. The attribute node exists to specify any of three node type: vertex node, relation node and attribute node. An attribute node can further specify other attribute nodes.

Table 1 shows the connectivity constraints among three node types. Two connections are prohibited: between vertex node and vertex node, and between link node and relation node, constraint imposed from their role in hypergraph. Table 1 is symmetrical on diagonal axis although the directionality of links depends on the context and what the network represents.

Figure 1 is an illustrative example of a hypernetwork representation. Attribute node linked to a vertex node specifies or defines the properties of the entity represented by the vertex node. Attribute node linked to another attribute node is homologous to the previous case, and it defines the properties of the quality or concept expressed by the other attribute

node. On the other hand, attribute node connected to a link node is the element absent in conventional knowledge representation models. This connection enables a detailed specification of relationship among vertex nodes. Note that the relationship treated here is N-ary which covers the binary relationship, the only relationship that conventional models can represent.

In this paper, we define knowledge as a structure consisting of terms and semantic relationships among terms. Thus two issues should be addressed: the extraction of technical terms and of semantic relationships among terms. The architecture of the knowledge structure is a hypernetwork, a whose representation power is higher than conventional knowledge representation models such as ER-model and semantic network.

Since the knowledge infrastructure is composed of terms, the integration of the knowledge of patients, doctors and researchers is executed by merging the terms used by these three types of people. Currently, two types of knowledge are being integrated: (1) Expert knowledge extracted from proceedings of the conference in the field of diabetes research; (2) Patients’ knowledge extracted from blogs writing about life with diabetes, often denoted as diabetic diary.

To accomplish this “knowledge merge”, we classified in two groups, (1) Patients and (2) Doctors and Researchers, where the first group has limited or no expert knowledge of the disease, and the second group has expert knowledge of the disease and background fields.

This paper analyzes the terms used in these two groups, necessary to build the knowledge infrastructure that serves people of both groups. Terms used by patients are extracted from blogs, and terms used by doctors/researchers are extracted from conference proceedings.

The target disease is diabetes. Here we denote the terms used by doctors/researchers as expert terms, and the terms used by patients as patient terms.

Expert terms are extracted from the proceedings of Japanese Diabetes Annual Conference between years 2007–2011 (JDS 2007; JDS 2008; JDS 2009; JDS 2010; JDS 2011). The title and abstract fields are extracted from the proceedings texts, totaling 9,335 presentation abstracts. The text data is then tagged with part of speech type using MeCab¹, and nouns are extracted. A total of 473,865 nouns were extracted, and duplicated entries were deleted. This list was manually checked to exclude meaningless words, resulting in 18,973 unique technical terms, or unique expert terms.

Patient terms, on the other hand, are extracted from 100 blogs written in Japanese by diabetes patients describing their dairy life related to their disease. The 100 blogs were selected from the top 100 Japanese blog sites in Google search results with query words “diabete” and “diary” (both words in Japanese). Each blog was manually checked to ensure the contents are actually about diabetes. The texts were manually extracted and nouns are extracted using similar procedure. Then the list of unique patient terms are cross-matched with unique expert terms.

Only 18.4% of unique expert terms matched with pa-

¹<http://code.google.com/p/mecab/>

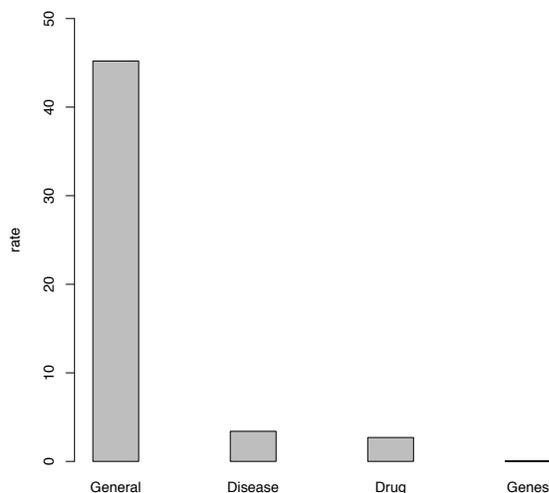


Figure 2: Crossmatch rate of expert terms with patient terms, individual values for general/other terms, disease names, drug names and substance/gene names.

tient terms. This low crossmatch rate indicate that a significant gap exists among expressions of patients and doctors/researchers.

For more detailed analysis, the expert terms extracted from proceedings are classified into (a) substance and gene names, (b) drug (medicine) names, (c) disease names, and (d) other terms including general nouns. The crossmatch rate of each class was relatively high for general and other nouns, very low for drug and disease names, and practically zero (less than 0.1%) for substance and gene names (Figure 2).

Word frequency is not considered in the analysis, because the knowledge infrastructure should possess all terms independent of their frequencies.

These results indicate that not only the language is different between patients and doctors, but a considerable gap exists between the viewpoints of patients and doctors. The viewpoint means which facet of the target topic is focused, treated or discussed. If no perspective difference existed, the corssmatch rate should have been approximately the same value for all four term classes (Figure 2). However, a clear disparity exists, where patients discusses very little about genes, opposite to the interest of doctors and researchers.

Therefore, the representation model of knowledge infrastructure should be capable to represent multiple viewpoints of a single concept. The hypernetwork model is able to incorporate multiple viewpoints, which is impossible with conventional models.

References

- Maeshiro, T., Nakayama, S., Maeshiro, M. 2011. Representation of Decision Making Process in Music Composition Based on Hypernetwork Model. In *Human Interface, Part I*, HCII 2011, LNCS 6771, 109-117. Springer-Verlag.
- Quillian, M.R. 1967. Word concepts: a theory and simulation of some basic semantic capabilities. *Behavioral Science*, 12(5): 410-430.
- Berge, C. 2001. *The Theory of Graphs*, Dover.
- Berge, C. 1989. *Hypergraphs: Combinatorics of Finite Sets*, North-Holland.
- JDS eds. 2007. *Proceedings of the 50th Annual Meeting of the Japan Diabetes Society*, Japan Diabetes Society.
- JDS eds. 2008. *Proceedings of the 51th Annual Meeting of the Japan Diabetes Society*, Japan Diabetes Society.
- JDS eds. 2009. *Proceedings of the 52th Annual Meeting of the Japan Diabetes Society*, Japan Diabetes Society.
- JDS eds. 2010. *Proceedings of the 53th Annual Meeting of the Japan Diabetes Society*, Japan Diabetes Society.
- JDS eds. 2011. *Proceedings of the 54th Annual Meeting of the Japan Diabetes Society*, Japan Diabetes Society.