FCAView Tab: A Concept-oriented View Generation Tool for Clinical Data Using Formal Concept Analysis

Guoqian Jiang1) Katsuhiko Ogasawara2) Naoki Nishimoto1) Akira Endoh1) Tsunetaro Sakurai1)
Department of Medical Informatics, Hokkaido University Graduate School of Medicine, Sapporo, Japan1)
Department of Health Science, Hokkaido University School of Medicine, Sapporo, Japan2)

Abstract: In the study, the potential role of formal concept analysis in concept-oriented view generation for clinical data is investigated. A concept-oriented view generation system for clinical data was developed as a Protege-2000 JAVA tab plug-in called FCAView Tab. A collection of discharge summaries in Japanese was mapped to Protege-2000 as a virtual patient record system. A natural language processing (NLP) module was developed to extract the relevant concepts from the records, based on using a Japanese standard dictionary of diagnostic terms and procedural terms released by MEDIS-DC. And a formal concept analysis (FCA) module was developed to generate the concept oriented views for the clinical data. The system could generate the views not only for the individual patient records, but also for the similar cases. The related issues and the future works were discussed.

1. Introduction

The computer-generated, concept-oriented views for clinical data can be used to reduce clinician information overload and improve the accuracy of clinical data retrieval [1]. However, the tasks for automating the generation of concept-oriented views are challenging. Formal concept analysis, a mathematical approach to data analysis based on the lattice theory, has been advocated to represent and process medical knowledge in different topics related to decision support systems [2]. In this study, the potential role of formal concept analysis in concept-oriented view generation for clinical data is investigated.

2. Materials

A collection of 386 Japanese discharge summaries obtained from Department of Cardiovascular Medicine of Hokkaido University Hospital was used.

3. System Construction

A concept-oriented view generation system for clinical data was developed as a Protege-2000 JAVA tab plug-in called FCAView Tab. Protege-2000 is an ontology development environment developed by Stanford Medical Informatics [3].

3.1 The virtual patient record system

A collection of Japanese discharge summaries represented by a XML file was mapped to Protege-2000 as a virtual patient record system. The items included File ID, Patient ID, Patient Age, Patient Sex, Patient Occupation, Admission Date, Discharge Date, Primary Diagnosis, Secondary Diagnosis, Complaints, Current History, Past History, Other History, Examination Records, Investigation Records, and Progress Records.

3.2 The natural language processing module

A natural language processing (NLP) module was developed to extract the relevant concepts from the virtual patient records, based on using a Japanese standard dictionary of diagnostic terms and procedural terms released by MEDIS-DC [4]. The construction of the module is described in our previous study [5].

3.3 The formal concept analysis module

A formal concept analysis (FCA) module was developed to generate the concept oriented views for the clinical data. An ontology description of FCA was made for the key concepts including formal attributes, formal objects and formal concept and their relationship. The Java API of an open source software Concept Explorer ver 1.2 was used to generate the concept oriented views for the clinical data [6].
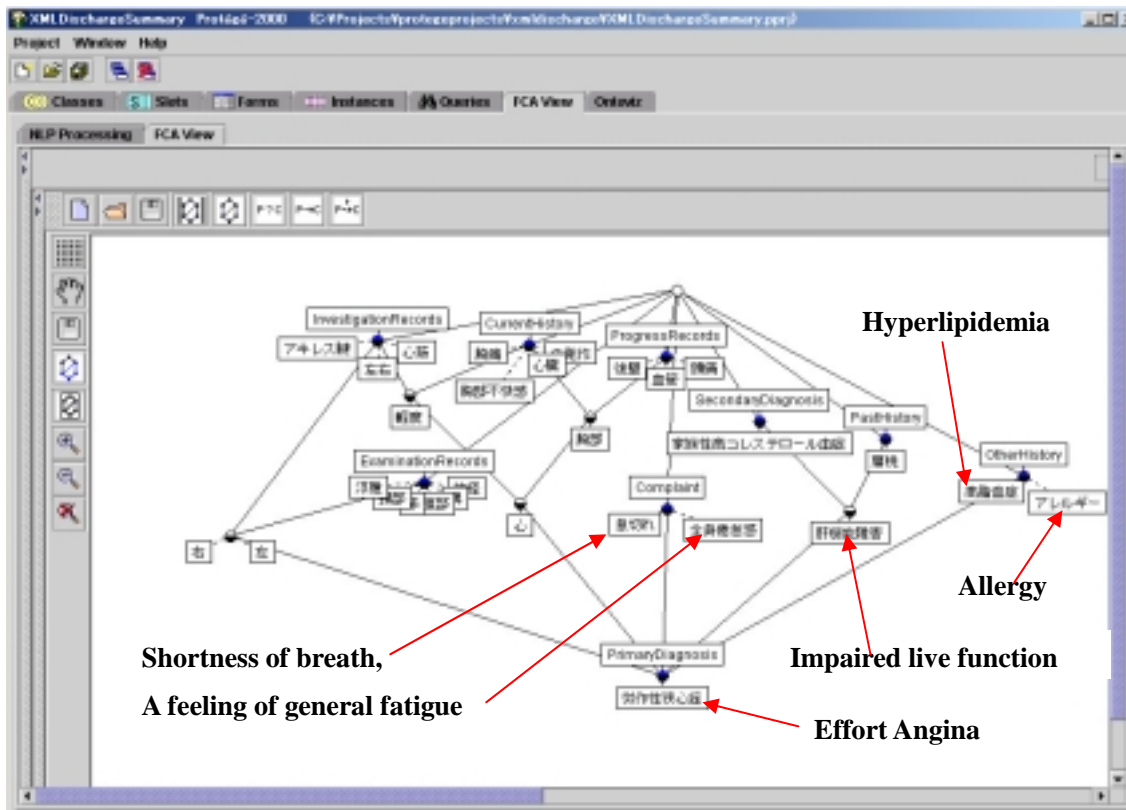
Fig. 1 An example view of a case with the primary diagnosis of effort angina (Red arrows indicate the English translations made by the authors)
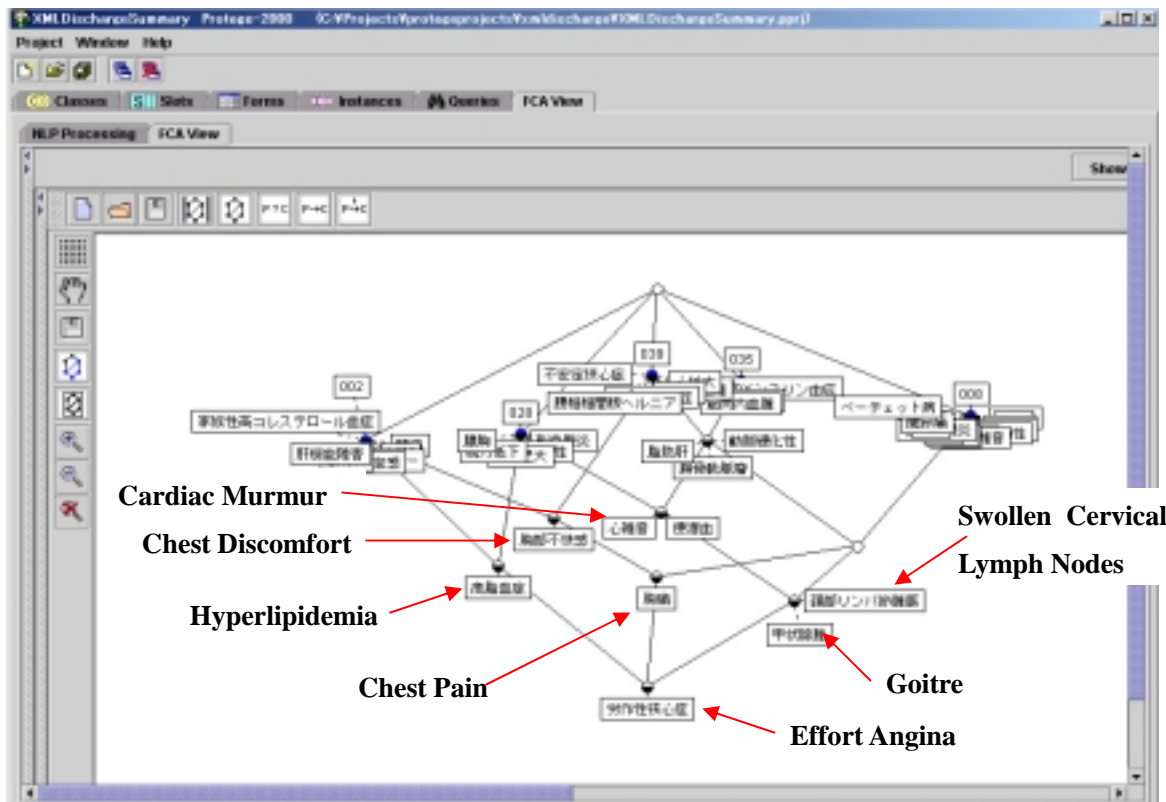

Fig. 2 An example view of ten cases with the primary diagnosis of effort angina (Red arrows indicate the English translations made by the authors)

4. Results

The system could generate the views for individual patient records. Fig.1 shows an example view of a case with the primary diagnosis of effort angina. Here the structure knowledge of the patient records was used as the formal objects and the medical concepts extracted from each part used as the formal attributes.

The system could also generate the views for the similar cases of a specific case. Fig. 2 shows an example view of ten cases with the same primary diagnosis of effort angina. Here the medical concepts extracted from each case were used as the formal attributes and the ten cases used as the formal objects.

5. Discussions

One of the features in our system is considered as automating concept-oriented view generation for clinical data by integrating the NLP, ontology-based knowledge base with the FCA technique, whereas most of systems in the past generated concept-oriented views manually and few systems did automatically [1].The most items of clinical data used in our system are textual and the concept-oriented views generated are based mainly on these textual data. From the clinical perspective, this kind of textual information contains the rich knowledge that could possibly benefit the clinicians to improve the patient care. Therefore, the key step here is to identify information relevant to the concepts of interest.

In this study, we constructed a NLP module to identify and extract the domain-specific concepts from textual patient records. In the NLP module, the MEDIS-based diagnosis terms and procedure terms are used as the main knowledge source. We have noticed that the knowledge source lacks of enough semantic relationship for the purpose of concept-oriented view generation. In the future this terminology issues could be possibly resolved by the ontology-based approach as demonstrated in our previous study [5].

In clinical domains, our previous study demonstrated that the FCA could be used to extract the semantic relationship between medical concepts [5].. Thus, the technique is possibly used to establish the semantic relationships among medical concepts and to compensate for the shortage of knowledge sources used in NLP module. Using the structural knowledge in an individual patient record, a lattice diagram could be produced automatically for concept-oriented view. In addition, the meta-knowledge of the similar cases was addressed a lot in the literatures related to the evidence-based clinical practice. Concept-oriented view for the similar cases identified the meta-knowledge and would possibly improve the evidence-based clinical practice of domain clinicians. The future works will focus on the generalization and evaluation of the system.

Considering the general use of the system, we developed a basic model of the FCAView Tab for generalization [7]. We assume that the instances (i.e. the formal objects) of a class in protégé-2000, together with its own slots (i.e. the formal attributes), could form a formal context that would interests the users. In particular, two kinds of formal context can be formed in the basic model: 1) the instances of a class with its slots in Boolean type; 2) the instances of a class with it slots in Multiple Instance type.

6. Acknowledgements

7. References

[1]Zeng Q, Cimino JJ, Zou KH. Providing concept-oriented views for clinical data using a knowledge-based system: an evaluation. J Am Med Inform Assoc. 2002 May-Jun;9(3):294-305.
[2]Schnabel M. Representing and processing medical knowledge using formal concept analysis. Methods Inf. Med. 41 (2002), pp. 160-167.
[3] http://protege.stanford.edu/index.html
[4] http://www.medis.or.jp/index.html
[5]Jiang G, Ogasawara K, Endoh A, Sakurai T. Context-based ontology building support in clinical domains using formal concept analysis. Int J Med Inform. 2003 Aug;71(1):71-81.
[6] http://sourceforge.net/projects/conexp
[7] http://info.med.hokudai.ac.jp/dept/gqjiang/fcaviewtab/fcaviewtab.html