Selected Studies from Current Medical Informatics Research in Japan

H. Tanaka
Tokyo Medical and Dental University, School of Biomedical Sciences, Tokyo, Japan

This is the second Japanese issue of MIM, subsequent to the first publication last year [1-9]. We are very thankful to Prof. Haux, who gave us the chance to publish a Japanese Special Topic in Methods of Information in Medicine (MIM) again this year. We invited submissions for this current issue from the Japanese medical informatics community. Among quite a few papers submitted, we have carefully selected six papers. It is often said that the medical informatics activities in Japan are difficult to grasp in its entirety because the Japanese researchers tend not to reveal their activities in the international journals. We wish this issue may contribute to giving MIM readers an overall perspective for current medical informatics activities in our country. We think the articles published in this issue reflect some representative aspects of the medical informatics researches in Japan, though not comprehensively covering its whole.

Like in other parts of the world, especially as in Europe and in North America, medical informatics research in Japan shares the common interests in many fields of medical informatics: we discuss the strategy to establish the nation-wide EHR system in Japan, make attempts to establish the interoperable standards among healthcare IT systems, develop the ontological basis of Japanese medical terminology, find ways to apply the recent communication technology to telemedicine or ubiquitous healthcare, try to bridge medical informatics and bioinformatics, and so on. But it sometimes occurs that in grappling with the research topics, we have to face challenges which are peculiar to our country: challenges that arise due to the characteristics of our healthcare systems, or those that occur due to the specificity of the Japanese language. Although all of the articles are engaged in studies to solve universal research challenges in medical informatics, some studies have to partly make certain effort to solve the problems peculiar to Japan. Such articles may be helpful to the readers who are facing the nation-specific problems described above.

For example, the first paper by Nishimoto et al. [10] studies the medical-text parsing algorithm for Japanese radiology reports. In applying national language processing to extract the clinical information from the medical text, the problem of delimiters is one of the difficulties we would face with Japanese; in English language, delimiters are clearly recognizable as words are separated by spaces, whereas in Japanese there are no spaces separating each word, making sentences just continuous sequence of Chinese and Japanese characters. With medical terms difficulties are made much worse as we must deal with long compound terms with many adjectives attributed to medical technicalities. In solving this problem, the authors investigated the transitional probability distribution of medical term boundaries between characters which was extracted from several hundred CT reports, thus improving the accuracies of parsing compound medical terms.

“Personalized medicine” based on individual genomic polymorphisms is also considered in Japan as a kind of goal which genomic medicine should pursue. In recent years, in order to attain the goal, genome-wide association study (GWAS) of human genome polymorphisms such as SNPs (single nucleotide polymorphisms) in relation to susceptibility of diseases has been conducted widely. The article by Toyabe et al. [11] aims to develop a new method to reduce the effort for analysis by statistical genetics applied to GWAS data. Their method consists of 1) creating the virtual genotype data of both “case” and “control” subjects from the database of the International HapMap Project by using bootstrap
simulation method with frequency of haplotypes being fixed, and 2) repeating virtual case-control studies and selecting the candidate SNPs. By this method, they succeeded to reduce the number of candidate SNPs which should be investigated for their relation to causes or susceptibility of diseases.

To describe the temporal course of diseases using Markov model has its long history of studies, but the article by Ishida et al. [12] develops the Markov model of the disease course of hepatic cellular carcinoma (HCC) which incorporates the effect of various therapies depending on the state of HCC with the subsequent subtree that might lead to recurrence. The authors used the retrospective data of 793 patients to calculate the monthly transient probabilities with good agreement to the prognosis data from nation-wide survey. They propose the cost-effective analysis of alternative therapies for HCC.

The article by Suzuki et al.[13] describes the study to extract index terms to disease recorded in hospital discharge summaries and select the DPC (Diagnostic Procedure Combination) code, which is the Japanese version of the DRG (diagnosis-related group). DPC classification is currently used as the basis of the inclusive evaluation system of hospitalization payment in approximately 400 hospitals in Japan. The authors used approximately 6000 patient cases to generate the document vector space according to the DPC. In addition, about 3000 cases, apart from the 6000 cases above, were used to verify the automatic DPC selection and 80 percent of cases matched the diagnosis of the DPC.

The article by Kawazoe and Ohe [14] describes the ontology-based study to extract the clinical knowledge for developing the decision support system. At first, they point out the difference of information model between the domain concept hierarchy and relational model used in the database of hospital information system (HIS) and also between the temporal-interval model and time-stamped data of the HIS database. To apply the clinical information to decision support systems, the relational and time-stamped data in HIS must be mapped onto the ontology-based concept models which are composed of clinical object model, clinical event model, and temporal model. After mapping to the ontology-based concept model, a rule-based inference system is implemented. In this system the rules to detect the adverse effect of drug are implemented. The system generates 615 alerts, where 346 were considered appropriate.

The article by Kimura et al. [15] describes high-speed clinical data retrieving system, from ten years of data of an operating hospital information system. The data were sent from HIS to HL7 v2.5 formatted retrieval system. To execute the search for the temporal relation quicker, the data structure is altered by moving the time stamp to the higher hierarchical position. This system showed sufficient high-speed retrieval.

When we selected the articles for publication, we took the precedence of technically sound research. Hence, many studies, especially those related with actual development of healthcare information system such as from HIS to nation-wide health care IT policies or regional healthcare system were dropped. Research in these fields would be collected on another occasion, as it would be also interesting for MIM readers to know the current situation of actual health information systems in Japan.

Anyhow, we wish this issue may be of use for the readers who might be interested in current research activity in Japan

References

Correspondence to:
Hinata Tanaka
School of Biomedical Sciences
Tokyo Medical and Dental University
1-5-45, Yushima, Bunkyo
Tokyo, Japan 113-8510
E-mail: tanaka@um.tmd.ac.jp