博士論文（要約）

A Framework of Mapping EHR Data to Disease Knowledge Presenting
Causal Chains of Abnormal States
– Chronic Kidney Disease Exemplar –

（疾患知識における異常状態と電子カルテデータとの自動マッピング
に関する研究
―慢性腎臓病を例として―）

馬　　驍珺

Precision medicine is an emerging approach to customization of disease interventions based on individual variability in genes, environment, and life style. Deep phenotyping defined as the precise and comprehensive analysis of the individual phenotypic abnormalities observed in a patient, which is expected to enable discovery of subclasses of a disease, has drawn great attention in support of precision medicine. This study proposes to present patients' data through disease ontology (defined as a causal chain of abnormal states) grasping the transitions of abnormal states, which accelerates the discovery of nuanced phenotypic abnormality serving deep phenotyping and disease progression management. For that purpose, automatic mapping between electronic health records (EHR) and disease ontology is indispensable. This study put forward to a framework of constructing a system of automatically extracting information from EHR, identifying presence of abnormal states, and mapping them to disease ontology. The framework consists of methodology of building mapping modules and a system composed of these modules which automatically maps patients' EHR data to disease ontology. This study first investigates EHR data sources from which information can be extracted for identification of abnormal states in ontologies of chronic diseases. We categorize data sources in EHR as clinical notes, imaging reports, laboratory data, treatment orderings, and demographics, which each require different mapping techniques. The proposed framework incorporates development of mapping modules from: (1) laboratory results of blood and urine tests; (2) medication prescriptions of oral medications and injections; (3) clinical texts of clinical notes including a great variety of documents (e.g., progress notes, discharge summaries, and nursery records), as well as imaging reports transcribed to clinical notes. We apply rule-based algorithms to estimating abnormal states from structured data of laboratory values and medication prescriptions, and propose a hybrid natural language processing (NLP) system combining rule-based and machine learning methods to identify abnormal states in unstructured clinical texts. The proposed system composed of multiple mapping modules gives a comprehensive

estimation on the presence of an abnormal state in a designated time window. Three types of imputations are implemented in our system to improve performance. One is along the timeline and the other two are based on causal chain structure of disease ontology. The imputations act on boosting certainty of evidences in EHR and treating data sparsity. This study takes chronic kidney disease (CKD) as exemplar. In this study, we have revised the CKD ontology and built a system of mapping EHR to the CKD ontology. The data used in this study are EHR data stored in SS-MIX2 standard and extended storage from The University of Tokyo Hospital generated during 2010 to 2017. The patient group for this study is selected from the patients registered to J-CKD-DB in 2014 from The University of Tokyo Hospital. Medical doctors have been involved in data annotation for evaluation. The experimental results demonstrate that the proposed framework is able to identify those abnormal states with high performance, which have little disagreement among doctors when annotation.