

Chapter XIV

Extracting Patient Case Profiles with Domain-Specific Semantic Categories

Yitao Zhang

The University of Sydney, Australia

Jon Patrick

The University of Sydney, Australia

ABSTRACT

The fast growing content of online articles of clinical case studies provides a useful source for extracting domain-specific knowledge for improving healthcare systems. However, current studies are more focused on the abstract of a published case study which contains little information about the detailed case profiles of a patient, such as symptoms and signs, and important laboratory test results of the patient from the diagnostic and treatment procedures. This paper proposes a novel category set to cover a wide variety of semantics in the description of clinical case studies which distinguishes each unique patient case. A manually annotated corpus consisting of over 5000 sentences from 75 journal articles of clinical case studies has been created. A sentence classification system which identifies 13 classes of clinically relevant content has been developed. A golden standard for assessing the automatic classifications has been established by manual annotation. A maximum entropy (MaxEnt) classifier is shown to produce better results than a Support Vector Machine (SVM) classifier on the corpus.

INTRODUCTION

The medical diagnosis and treatment of a patient is a complex procedure which requires relevant

knowledge and clinical experience of illnesses. In order to distinguish different illnesses that show similar signs and symptoms, or to decide the best available option for treatment of a medi-

cal condition, physicians need to have adequate observations of similar patient cases, either from their own previous medical practices, or some external resources, such as the knowledge of more experienced colleagues, and the medical literature on the latest progress in the field.

Clinical case reporting therefore plays an important role in both educating young physicians in the practice of medicine, and sharing clinical experience and exceptional findings among physicians (Jenicek, M., 2001). There are two types of clinical case reports, namely routine patient case reports and clinical case studies. Routine case reports are raw patient records of the diagnostic and treatment procedure and provide information which is necessary for the continuity of patient care, such as progress notes, discharge summaries and pathology reports. Clinical case studies, on the other hand, report rare and abnormal patient cases that are considered as of significant scientific value to the field. They are reported by clinicians, usually in the form of formal journal articles in medical press.

Narrative patient records produced by nurses and physicians everyday in hospitals and clinics, provide first-hand the richest information about the progress of patients. However, the confidentiality of personal records has always been a concern which prevents the research community having access to enough data to develop useful learning systems comparable to human performance. Confidential information includes names of patients and physicians, dates, and geographic clues which are required to be anonymized before any raw patient data can be released to the public. Moreover, the anonymisation task often requires human annotators to manually check every single patient record to satisfy certain laws and ethical guidelines specified by governments. For instance, the 2007 Computational Medicine Challenge had to use human annotators to review all of the 4,055 raw patient records and to remove nearly half of them from the final gold-standard corpus to meet United States HIPAA standards.

With the emergence of publicly available on-line knowledge bases such as MEDLINE/PubMed and BMC Central, clinicians now have access to a large number of full-text journal articles of clinical case studies. Each case study records a detailed discussion of the patient's abnormal signs and symptoms, or novel conditions which are considered as report-worthy. While all the sensitive privacy information has been carefully removed from the text, a clinical case study still contains rich information about patient case profiles, such as patient demographics, signs and symptoms, laboratory test readings and interpretations, and treatments and subsequent outcomes for patients. This patient profile information is key in answering two fundamental questions dominating the daily practice of physicians: (1) Given the case profile of a patient, what is the best explanation or diagnosis of the condition? (2) Given the specified circumstances of a patient, what is the best treatment available? By exploring clinical case studies with similar patient profiles, physicians can learn, and therefore improve their practices of medicine, from the successes or failures of their peers.

However, the amount of data added into on-line biomedical databases each year is so large that clinicians cannot keep up to date with the new discoveries and knowledge in the field. For instance, there are 500,000 new journal articles added into the PubMed/MEDLINE database each year (Mitchell, J. A., Aronson, A. R., & Mork, J. G., 2003). An information need therefore arises for building intelligent information retrieval systems which incorporate domain-knowledge to find the right patient cases to support clinical decision making in a timely manner. Most efforts so far in biomedical text-mining have been made to exploit only abstracts of clinical case studies, which is still a small proportion of the fully available text (Cohen, A. M., & Hersh, W. R., 2003). Important clinical information, such as laboratory tests and readings, signs and symptoms of a patient during the diagnosis and treatment, and health profile of

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/extracting-patient-case-profiles-domain/23065

Related Content

An Overview of Graph Indexing and Querying Techniques

Sherif Sakrand Ghazi Al-Naymat (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 222-239).

www.irma-international.org/chapter/overview-graph-indexing-querying-techniques/76064

Graphical Analysis and Visualization Tools for Protein Interaction Networks

Sirisha Gollapudi, Alex Marshall, Daniel Zadikand Charlie Hodgman (2009). *Biological Data Mining in Protein Interaction Networks* (pp. 286-311).

www.irma-international.org/chapter/graphical-analysis-visualization-tools-protein/5570

RETRACTED: Human Creation and the Role of the Family: Theological Theory in Bioinformatics

Isahak Poghosyan, Tatul Manaseryanand Laura Aghajanyan (2022). *International Journal of Applied Research in Bioinformatics* (pp. 1-7).

www.irma-international.org/article/retracted-human-creation-and-the-role-of-the-family/282689

Bioinformatics Methods for Studying MicroRNA and ARE-Mediated Regulation of Post-Transcriptional Gene Expression

Richipal Singh Bindra, Jason T. L. Wangand Paramjeet Singh Bagga (2010). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 97-112).

www.irma-international.org/article/bioinformatics-methods-studying-microrna-mediated/47098

Similarity Searching of Medical Image Data in Distributed Systems: Facilitating Telemedicine Applications

Amalia Charisi, Panagiotis Korvesisand Vasileios Megalooikonomou (2013). *Methods, Models, and Computation for Medical Informatics* (pp. 58-77).

www.irma-international.org/chapter/similarity-searching-medical-image-data/73071