Volume : 1 | Issue : 2 | February 2012 ISSN : 2250 - 1991

Research Paper Engineering

A ML Approach For Identifying Semantic Relations In Short Texts That Exist Between Diseases And Treatments

* Sowmyadevi K **Reha S *** Malini A
**** Mohana Priya.G

* Computer Science Engineering Department, Anna University

** Mainframe Department, Anna University

*** Information Technology Department, Anna University

**** 4Computer Science Engineering Department, Anna University, Coimbatore

ABSTRACT

The Machine Learning (ML) field has gained its momentum in almost any domain of research and just recently has become a reliable tool in the medical domain. The empirical domain of automatic learning is used in tasks such as medical decision support, medical imaging, protein-protein interaction, extraction of medical knowledge, and for overall patient management care. ML is envisioned as a tool by which computer-based systems can be integrated in the healthcare field in order to get a better, more efficient medical care. This paper describes a ML-based methodology for building an application that is capable of identifying and disseminating healthcare information. It extracts sentences from published medical papers that mention diseases and treatments, and identifies semantic relations that exist between diseases and treatments. Our evaluation results for these tasks show that the proposed methodology obtains reliable outcomes that could be integrated in an application to be used in the medical care domain. The potential value of this paper stands in the ML settings that we propose and in the fact that we outperform previous results on the same data set.

Keywords : ML Approach, Diseases, Treatment

Introduction

People care deeply about their health and want to be, now more than ever, in charge of their health and healthcare. Life is more hectic than has ever been, the medicine that is practiced today is an Evidence-Based Medicine (hereafter EBM) in which medical expertise is not only based on years of practice but on the latest discoveries as well. Tools that can help us manage and better keep track of our health such as Google Health1 and Microsoft HealthVault2 are reasons and facts that make people more powerful when it comes to healthcare knowledge and management. The traditional healthcare system is also becoming one that embraces the Internet and the electronic world. Electronic Health Records (hereafter EHR) are becoming the standard in the healthcare domain Researches and studies show that the potential benefits of having an system are3EHR:Health information recording and clinical data repositoriesimmediate access to patient diagnoses, allergies and lab test results that enable better and time-efficient medical decisions; Medication managementrapid access to information regarding potential adverse drug reactions, immunizations, supplies, etc; - Decision supportthe ability to capture and use quality medical data for decisions in the workflow of healthcare.

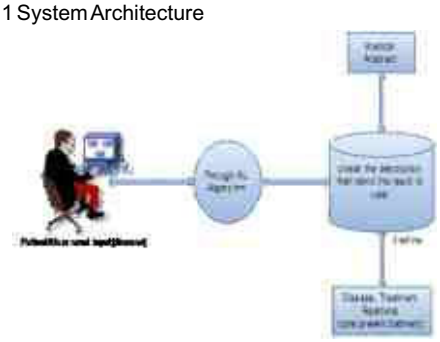
Architecture

In order to embrace the views that the EHR system has, we need better, faster, and more reliable access to information. In the medical domain, the richest and most used source of information is Medline, database of extensive life science published articles.

All research discoveries come and enter the repository at high rate (Hunter and Cohen, making the process of identifying and disseminating reliable information a very difficult task. One task is automatically identifying Sentences published in medical abstracts (Medline) as containing or not information about disease Treatments and automatically identifying semantic relations that exist between disease and Treatments. Use Hidden Markov Models and maximum entropy models to perform both the task of entity recognition and the relation discrimination.

Their representation techniques are based on words in context, part of speech information, phrases, and a medical lexical ontology Mesh terms. People want Fast access to reliable information and in a manner that is suitable to their habits and workflow. There is no reliable information.

Fig. 1 System Architecture



Our proposed research paper describes a ML-based methodology for building an application that is capable of identifying and disseminating healthcare information. The ML settings that we propose and in the fact that we out perform previous results on the same data set.

A. A Shortest Path Dependency Kernel For Relation Extraction

Present a novel approach to relation extraction, based on the observation that the information required to assert a relationship between two named entities in the same sentence is typically captured by the shortest path between the two entities in the dependency graph. Experiments on extracting top-level relations from the ACE (Automated Content Extraction) newspaper corpus show that the new shortest path dependency kernel outperforms a recent approach based on dependency tree kernels.

B. Textual Information For Predicting Functional Properties Of The Genes

In this paper is focused on determining which proteins affect the activity of Aryl Hydrocarbon Receptor (AHR) system when learning a model that can accurately predict its activity when single genes are knocked out. Experiments with results are presented when models are trained on a single source of information: abstracts from Medline that talk about the genes involved in the experiments. The results suggest that AdaBoost classifier with a binary bag-of-words representation obtains significantly better results.

C. Detecting influenza epidemics using search engine query data

Epidemics of seasonal influenza are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year1. In addition to seasonal influenza, a new strain of influenza virus against which no prior immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities2. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza3,4. One way to improve early detection is to monitor health-seeking behavior in the form of online web search queries, which are submitted by millions of users around the world each day. Here we present a method of analyzing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to utilize search queries to detect influenza epidemics in areas with a large population of web search users.

D. Protein Structures and Information Extraction From Biological Texts: The PASTA System

The rapid increase in volume of protein structure literature means useful information may be hidden or lost in the published literature and the process of finding relevant material, sometimes the rate-determining factor in new research, may be arduous and slow.

E. Learning to Extract Relations from MEDLINE

Information in text form remains a greatly underutilized resource in biomedical applications. We have begun a research e ort aimed at learning routines for automatically mapping information from biomedical text sources, such as MEDLINE, into structured rep resonations, such as knowledge bases. We describe our application, two learning methods that we have applied to this task, and our initial experiments in learning such information-extraction routines. We also present an approach to decreasing the cost of learning information-extraction routines by learning from \weakly" labeled training data.

Implementation Details

A ML-based methodology for building an application that is capable of identifying and disseminating healthcare information.

A. Clinical Decision Support System

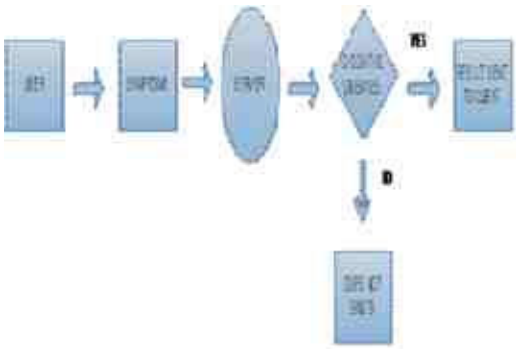
Clinical decision support system (CDSS or CDS) is an interactive decision support system (DSS) Computer Software, which is designed to assist physicians and other health professionals with decision making tasks, as determining diagnosis of patient data. A working definition has been proposed by Dr. Robert Hayward of the Centre for Health Evidence; "Clinical Decision Support systems link health observations with health knowledge to influence health choices by clinicians for improved health care".

The main purpose of modern CDSS is to assist clinicians at the point of care. This means that a clinician would interact with a CDSS to help determine diagnosis, analysis, etc. of patient data. Previous theories of CDSS were to use the CDSS to literally make decisions for the clinician. The clinician would input the information and wait for the CDSS to output the “right” choice and the clinician would simply act on that output. The new methodology of using CDSS to assist forces the clinician to interact with the CDSS utilizing both the clinician's knowledge and the CDSS to make a better analysis of the patients data than either human or CDSS could make on their own. Typically the CDSS would make suggestions of outputs or a set of outputs for the clinician to look through and the clinician officially picks useful information and removes erroneous CDSS suggestions.

An example of how a CDSS might be used by a clinician comes from the subset of CDSS, DDSS (Diagnosis Decision Support Systems). A DDSS would take the patients data and propose a set of appropriate diagnoses. The doctor then takes the output of the DDSS and figures out which diagnoses are relevant and which are not.

Another important classification of a CDSS is based on the timing of its use. The doctor uses these systems at point of care to help them as they are dealing with a patient, with the timing of use as either pre-diagnosis, during diagnoses, or post diagnoses. Pre-diagnoses CDSS systems are used to help the physician prepare the diagnoses. CDSS used during diagnoses help review and filter the physician's preliminary diagnostic choices to improve their final results. And post-diagnoses CDSS systems are used to mine data to derive connections between patients and their past medical history and clinical research to predict future events.

Fig. 2 Prediction

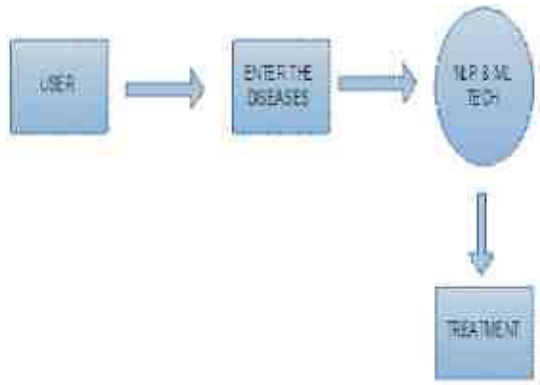


A. Extraction Of Medical Knowledge

Modern clinical practice emphasizes medical research evidence, physicians' clinical expertise, as well as patients' individual values (e.g. preference, concerns, and expectations). Decision analysis is an effective method for structuring clinical problems, encoding objective evidences, representing clinicians' subjective judgments and expressing patients' preferences to derive optimizing solutions in diagnostic, therapeutic, and prognostic management. The main task in decision analysis is to construct decision models, which are mathematical frameworks with graphical representations. This work focuses on the construction of Influence Diagrams (IDs), a powerful tool in structuring complex problems and facilitating communication between the physicians and decision analysts.

Decision model construction is a knowledge intensive task, involving one or more decision analysts working closely with one or more domain experts to elicit the relevant structural and numerical parameters of the decision models. It is usually an arduous task to process and integrate all the knowledge needed for model construction. In addition, as biomedical knowledge is advancing rapidly, the domain experts may not always be up-to-date on all the latest advancements in a particular field during the model construction process. Hence it is desirable to automatically derive a summary of the useful relations among the medical concepts relevant to a decision problem, which can in turn help build a proper context of the knowledge required for full decision model construction.

Fig. 3 Extraction of medical knowledge



A. EHR Electronic Health Record

An electronic health record (EHR) (also electronic patient record (EPR) or computerized patient record) is an evolving concept defined as a systematic collection of electronic health information about individual patients or populations. It is a record in digital format that is capable of being shared across different health care settings, by being embedded in network-connected enterprise-wide information systems. Such records may include a whole range of data in comprehensive or summary form, including demographics, medical history, medication and allergies, immunization status, laboratory test results, radiology images, vital signs, personal stats like age and weight, and billing information.

Its purpose can be understood as a complete record of patient encounters that allows the automation and streamlining of the workflow in health care settings and increases safety through evidence-based decision support, quality management, and outcomes reporting.[

1) Improve quality of care

The implementation of electronic health records (EHR) can help lessen patient sufferance due to medical errors and the inability of analysts to assess quality. Information Technology is being used today to automate day-to-day processes, thus helping to reduce administration costs which then in turn can free up time and money for patient care.

EHR systems can help reduce medical errors by providing healthcare workers with decision support.[Fast access to medical literature and current best practices in medicine are hypothesized to enable proliferation of ongoing improvements in healthcare efficacy. Improved usage of EHR is achieved if the presentation on screen or on paper is not just longitudinal, but hierarchically ordered and layered.

2) Promote evidence-based medicine

EHRs provide access to unprecedented amounts of clinical data for research that can accelerate the level of knowledge of effective medical practices. Realistically, these benefits may only be realized if the EHR systems are interoperable and wide spread (for example, national or regional level) so that various systems can easily share information. Also, to avoid failures that can cause injury to the patient and violations to privacy, the best practices in software engineering and medical informatics must be deployed.

Fig. 4 Information Pack

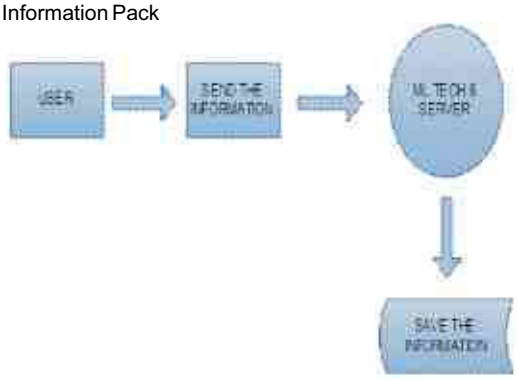
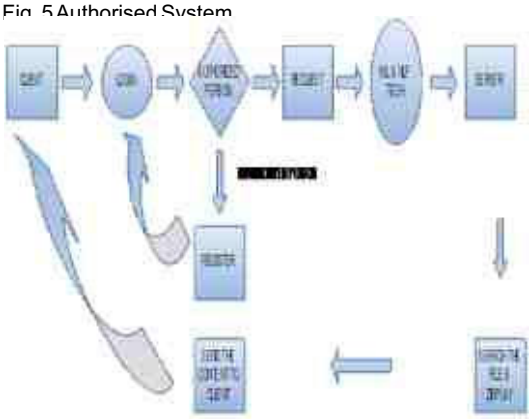


Fig. 5 Authorised System



Conclusions

The work that present in this paper is focused on two tasks: automatically identifying sentences published in medical abstracts (Medline) as containing or not information about diseases and treatments, and automatically identifying semantic relations that exist between diseases and treatments, as expressed in these texts. The second task is focused on three semantic relations: Cure, Prevent, and Side Effect. Our objective for this work is to show what Natural Language Processing (NLP) and Machine Learning (ML) techniques what representation of information and what classification algorithms are suitable to use for identifying and classifying relevant medical information in short texts. We acknowledge the fact that tools capable of identifying reliable information in the medical domain stand as building blocks for a healthcare system that is up-to-date with the latest discoveries.

REFERENCES

S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998. | J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61. | S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," IEEE Electron Device Lett., vol. 20, pp. 569571, Nov. 1999. | M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109. | R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997. | (2002) The IEEE website. [Online]. Available: http://www.ieee.org/ | M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/ | FLEXChip Signal Processor (MC68175/D), Motorola, 1996. | "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland. | A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999. | J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999. |Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.