International Journal of Science Engineering and Advance Technology

Medical Decision Support Using Machine Learning And Natural Language Processing

Bulusu Krishna Sashidhar, Srinivasa Babji Josyula
[1]M.Tech on Computer Science with the specialization In SE
[2]Assistant Professor in Department of computer Science
& System Engineering Andhra University College of Engineering (Autonomous)

ABSTRACT

The Machine Learning (ML) field has got much attention among the researchers in almost many fields of science and engineering because of it its vast application. The common application areas are medical decision support, medical imaging, protein-protein interaction, extraction of medical knowledge, and for overall patient management care. ML is used as a tool to introduce computer-based systems in the healthcare field in order to get a better, more efficient medical care. In this work Machine Learning and Natural language Processing are used for extracting knowledge from published medical papers. It extracts the sentences that mention diseases and treatments and identifies therelationship between them.

INTRODUCTION

People care much about their health and want to be, now more than ever, in charge of their health and healthcare. Tools that can help us manage and better keep track of our health such as Google Health and Microsoft HealthVault 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 are becoming the standard in the healthcare domain.

The objective is to speed up the interaction between doctors and patients and coordinate various activities that take place in atypical hospital and to work with ML and NLP is the task of identifying and disseminating reliable healthcare information becomes easy and beneficial for the people. A hierarchical approach is used for performing the two tasks: The first is to identify and eliminate uninformative sentences and then second is to classify the rest of the sentences by the relation of interest. By this a substantial improvement is shown getting the information.

2. LEARNING TO EXTRACT RELATIONS FROM MEDLINE

Information in text form remains a greatly underutilized resource in biomedical applications. We have begun a research effort aimed at learning routines for automatically mapping information from biomedical text sources, such as MEDLINE, o structured representations, 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.

The MEDLINE database is a rich source of information for the biomedical sciences, providing bibliographic information and abstracts for more than nine million articles. A fundamental limitation of MEDLINE and similar sources, however, is that the information they contain is not represented in structured format, but instead in natural language text. The goal of our research is to develop methods that can inexpensively and accurately map information in scientific text sources, such as MEDLINE, to a structured representation, such as a knowledge base or a database. Toward this end, we are investigating methods for automatically extracting key facts from scientific texts.

There are three aspects of our work that we think will be of particular interest to the *machine learning for information extraction* community. First, our application domain is novel and challenging. Second, we investigate an approach to decreasing the cost of learning information-extraction routines by learning from "weakly" labeled training data. Weakly labeled instances consist not of precisely marked up documents, but instead consist of facts to be extracted

along with documents that may assert the facts. Third, we employ a learning method we have recently developed that incorporates statistical predicate invention into a relational learner. The system we are developing is motivated by several different types of tasks that we believe would greatly benefit from the ability to extracted structured information from text.

Especially compelling application of our system is its potential application to scientific discovery. The articles in MEDLINE scribe a vast web of relationships among the genes, proteins, pathways, tissues and diseases of various systems and organisms of interest. Moreover, each article describes only a small piece of this web. The work of Swanson and Smalheiser (1997) has demonstrated that significant but previously unknown relationships among entities (e.g., magnesium and migraine headaches) can be discovered by automatically eliciting this information from the literature. Swanson's algorithm detects relationships among objects simply by considering the statistics of word co-occurrences in article titles.

2.1 TO IDENTIFY DISEASE TREATMENT RELATIONSHIPS IN SHORT TEXT USING MACHINE LEARNING & NATURAL LANGUAGE PROCESSING

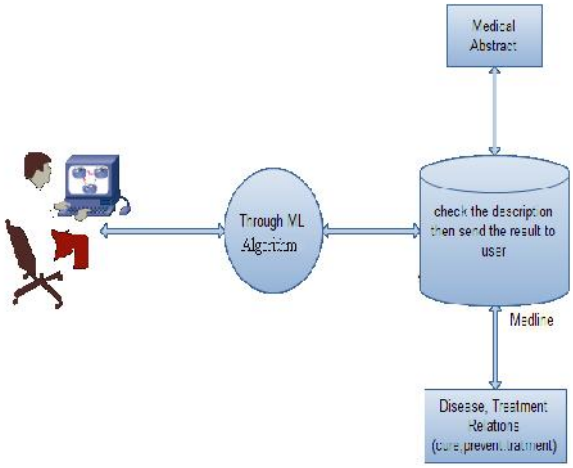
Due to advancements in medical domain automatic learning has gained popularity in the fields of medical decision support, extraction of medical knowledge and complete health management. Using Machine Learning (hereafter, ML) and Natural language Processing (hereafter, NLP) we can make the healthcare field more efficient and reliable. This paper describes how ML and NLP can be used for extracting knowledge from published medical papers. It extracts the sentences that mention diseases and treatments and identifies the relationship between them.

People are more concerned about their health than ever before. In spite of their busy schedules they want each and everything to go in a good flow. People want Fast access to reliable information and in a manner that is suitable to their habits and workflow. Medical field has grown to such an extent that the people practicing medicine should not only have experience but also information about latest discoveries. Electronic Health Record (hereafter, EHR) is becoming a standard in healthcare domain. Websites such as Google Health and Microsoft Health Vault make people to care deeply about their health.

EHR has the following benefits:

1. Rapid access to information that is focused on certain topics such as immunizations, drugs etc.
2. To have quality medical data for taking proper medical decisions. For this purpose we need a better, more efficient and reliable access to information.

According to researches people are searching the web in order to be informed regularly about their health. In medical domain the most used source of information is MEDLINE. MEDLINE is database where all the research discoveries come and enter at a high rate. Due to the busy schedules the experts don't get time to read millions of articles therefore there is a need to build a tool that will suffice the purpose.



Architecture of the system

3. SYSTEM DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- The dialog to guide the operating personnel in providing input.

- Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

3.1. OUTPUT

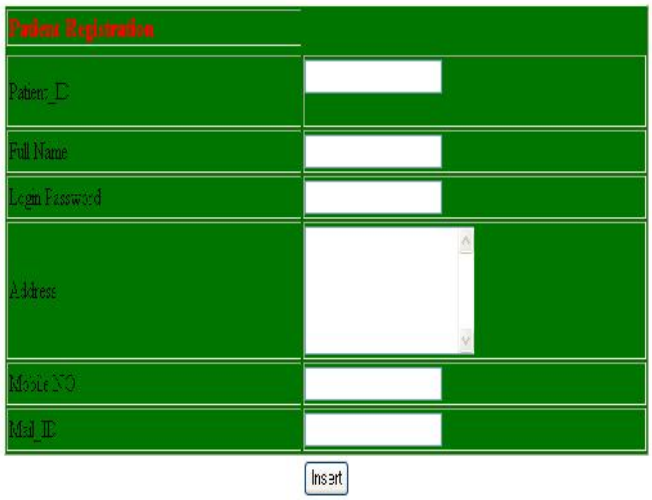
A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

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. We focus on diseases and treatment information, and the relation that exists between these two entities. Our interests are in line with the tendency of having a personalized medicine, one in which each patient has its medical care tailored to its needs. It is not enough to read and know only about one study that states that a treatment is beneficial for a certain disease. Healthcare providers need to be up-to-date with all new discoveries about a certain treatment, in order to identify if it might have side effects for certain types of patients.



A sample for data insertion

Result displays the disease details

4. CONCLUSION

This approach is very useful for everyone as it gives information only of the area of interest. The task is divided into two tasks the first task that we tackle in this paper is a task that has applications in information retrieval, information extraction, and text summarization. Our work shows that the best results are obtained when the classifier is not overwhelmed by sentences that are not related to the task.

5. REFERENCES

[1]Oana Frunza, Diana Ink pen, and Thomas Tran " A Machine Learning Approach for Identifying Disease-Treatment Relations in Short Texts" vol. 23, 2011.

[2] B. Rosario and M.A. Hearst, “Semantic Relations in Bioscience Text,” Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics, vol. 430, 2004.

[3] M. Craven, “Learning to Extract Relations from Medline,” Proc. Assoc. for the Advancement of Artificial Intelligence, 1999.

[4]S. Ray and M. Craven, “Representing Sentence Structure in Hidden Markov Models for Information Extraction,” Proc. Int’l Joint Conf. Artificial Intelligence (IJCAI ’01), 2001.

[5] P. Srinivasan and T. Rindflesch, “Exploring Text Mining from Medline,”Proc. Am. Medical Informatics Assoc. (AMIA) Symp. 2002.

[6] R. Bunescu, R. Mooney, Y. Weiss, B.Scho¨ lkopf, and J. Platt,“Subsequence Kernels for Relation Extraction,” Advances in Neural Information Processing Systems, vol.18, pp. 171-178, 20.

[7] Frunza, Oana, Diana Inkpen, and Thomas Tran. "A machine learning approach for identifying disease-treatment relations in short texts." Knowledge and Data Engineering, IEEE Transactions on 23.6 (2011): 801-814.