Natural language processing framework to assist in the evaluation of adherence to clinical guidelines

Sushmitha Regulapati
West Virginia University
Natural Language Processing Framework to assist in the Evaluation of Adherence to Clinical Guidelines

Sushmitha Regulapati

Thesis submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Computer Science

Dr. Juggy V.Jagannathan, Ph.D., Chair
Dr. Arun Ross, Ph.D.
Dr. Yenumula V.Reddy, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2007

Keywords: clinical guidelines, guideline representation models, natural language processing, healthcare, electronic health records, congestive heart failure (CHF)
ABSTRACT

Natural Language Processing Framework to assist in the Evaluation of Adherence to Clinical Guidelines

Sushmitha Regulapati

Congestive Heart Failure (CHF) is one of the leading causes associated with mortality and morbidity amongst the hospitalized patients and accounts to a major portion of health care costs in USA. Milliman Care Guidelines for CHF is used as a basis to develop a flowchart for representing the guidelines. Using Natural Language Processing techniques, a collection of sample discharge summaries from WVU university hospital are verified against the flowchart to check if they followed the Clinical Guidelines for CHF. The results suggest that it is feasible to evaluate clinical records against the clinical guidelines by representing the guidelines as a set of keywords. The framework developed is extensible can be applied to multiple guidelines.
Acknowledgements

Many people contributed to the successful completion of this thesis, most notable Dr. Juggy V. Jagannathan who has been an exceptional advisor and mentor through out my graduate career. He encouraged me all the time and was ready to extend a helping hand whenever needed. I have greatly benefited from his help and guidance through out my Master’s program. This work is done for Medquist, Inc.

I would like to acknowledge my committee members Dr. Yenumula V. Reddy and Dr. Arun Ross for taking time from their busy schedules for my thesis. I would also like to thank John Ninomiya, PHD, MSc. Director, Epidemiology and Health Services research Milliman Care Guidelines, LLC, a Milliman Company and also Dr. Charles Mullet of WVUH.

Finally I am most grateful to my family and friends, whose love and support made everything possible.
Table of Contents

Title Page…………………………………………………………………………………… i
Abstract……………………………………………………………………………………… ii
Acknowledgements………………………………………………………………………… iii
Table of Contents……………………………………………………………………………… iv
List of Figures………………………………………………………………………………… vi
List of Table…………………………………………………………………………………. vii
Chapter 1 Introduction ………………………………………………………………………… 1
1.1 Introduction ………………………………………………………………………………… 1
1.2 Clinical Care Guidelines …………………………………………………………………… 2
1.2.1 Introduction ………………………………………………………………………………… 2
1.2.2 Necessity for Clinical Guidelines ……………………………………………………… 2
1.2.3 Benefits of Clinical Guidelines ………………………………………………………… 3
1.2.4 Developing Clinical Guidelines ……………………………………………………… 4
1.2.5 Computerized Guidelines ……………………………………………………………… 5
1.3 Problem Statement ………………………………………………………………………… 5
1.4 Organization ………………………………………………………………………………… 6
Chapter 2 Literature Survey ………………………………………………………………… 7
2.1 Clinical Practice Guideline Representation Models …………………………………… 7
2.1.1 Introduction ………………………………………………………………………………… 7
2.1.2 Developing Computer Interpretable Guidelines from text ………………………… 7
2.1.3 Requirements of a Guideline Model ………………………………………………….. 8
2.1.4 Guideline Representation Models and Tools …………………………………………. 9
2.1.4.1 GLIF …………………………………………………………………………………... 10
2.1.4.2 Arden Syntax ………………………………………………………………………… 10
2.1.4.3 Asbru ………………………………………………………………………………… 11
2.1.4.4 EON …………………………………………………………………………………… 11
2.1.4.5 GEM ………………………………………………………………………………….. 12
2.2 Natural Language Processing …………………………………………………………….. 12
2.2.1 Introduction ………………………………………………………………………………. 12
2.2.2 NLP in healthcare ……………………………………………………………………… 13
2.2.2.1 Introduction ………………………………………………………………………… 13
2.2.2.2 MLP …………………………………………………………………………………. 14
2.2.2.3 NLP systems developed in healthcare …………………………………………... 14
2.2.2.3.1 MedLEE ………………………………………………………………………….. 14
2.2.2.3.2 MENELAS ……………………………………………………………………….. 15
2.2.2.3.3 DAD systems ……………………………………………………………………… 15
2.2.2.4 Natural Language Processing and clinical care guidelines ……………………. 16
Chapter 3 Design Approach ………………………………………………………………… 18
3.1 Design ……………………………………………………………………………………. 18
3.1.1 The Framework ………………………………………………………………………... 18
3.1.1.1 Research Corpus ………………………………………………………………… 18
3.1.1.2 Congestive Heart Failure ………………………………………………………… 19
3.1.1.3 Clinical Guidelines for CHF …………………………………………………….. 21
3.1.1.4 Flowchart Representation of the Guidelines …………………………………… 22
3.1.1.5 XML representation of the Guidelines ………………………………………… 24
List of Figures

Figure 1. The Framework 18
Figure 2. A sample discharge summary 19
Figure 3. Screenshot of Milliman Care Guidelines for CHF 22
Figure 4. Flowchart Representation of Clinical Guidelines for CHF 23
Figure 5. XML Representation of Guidelines 24
Figure 6. Process used for analysis 27
Figure 7. Evaluation results 30
List of Tables

Table 1. Result Analysis 28
Table 2. Results 29
Chapter 1

Introduction

1.1 Introduction

In response to the IOM report ‘To Err Is Human: Building a Safer Health System’ (Institute of Medicine, 2000) call to reduce errors in Health Care Systems focus on achieving patient safety as a standard for care has increased. [1] Patient Safety is prevention of harm to patients. One of the recommendations of the IOM report on ‘Patient Safety a new Standard for Care’ (Institute of Medicine, 2003) is that all health care organizations should provide immediate access to complete patient information and decision support tools (e.g. alerts, reminders) and a better management of health information would help achieve this. Electronic Health Record (EHR) is an electronic representation of health information that captures patient data and interacts with decision support applications in order to improve safety and quality of care. Data elements used in these records and applications have to be represented in a standard way i.e. standardized data interchange formats, terminologies and knowledge representation is essential.

Agency for Healthcare Research and Quality (AHRQ), an agency that devotes its attention to improving quality in healthcare, is supporting the use of specifications such as HL7 Clinical Document Architecture (CDA) and implementation guides in order to enhance interoperability. The agency should also consider the standards mandated by the Health Insurance Portability and Accountability Act to identify areas of overlap and gaps in terminologies. To standardize knowledge representation, AHRQ in collaboration with the National Institute of Health, the Food and Drug Administration and with support from
National Library of Medicine should support development of a generic guideline representation model and represent them in a computer-executable format.

1.2 Clinical Care Guidelines:

*Health care guidelines are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances"* (Institute of Medicine, 1992).

1.2.1 Introduction:

A guideline is a process specification useful in improving quality of care by bringing together the best external evidence and other knowledge necessary for decision-making about a specific health problem. The National Guideline Clearing House (NGC) organization lists the various guideline developer organizations (in NGC Browse). Guidelines may be developed by public or private organizations, institutions or expert panels. Produced by the Agency of Healthcare Research and Quality (AHRQ) in partnership with the American Medical Association (AMA) and the American Association of Health Plans (AAHP) foundation, it is a publicly available database of evidence-based clinical practice guidelines and related documents.

The audiences of clinical practice guidelines vary from individual physicians and other clinicians (for decision-making and patient counseling) and health care organizations to educational institutions, state and local governments (for quality assurance).

1.2.2 Necessity for Clinical Guidelines:
The clinical guidelines offer concise instructions regarding the various diagnostic and screening tests to be ordered in order to evaluate a patient for a specific disease and also offer instructions on how long patients should stay in a hospital for a specific disease. They also instruct on the discharge medications and procedures to be done during the hospitalization for a specific diagnosis. The necessity for clinical guidelines came from the rising health care costs and from the expectations of the patients for quality care irrespective of place, physician and the facility at which they are being treated. Physicians, policy makers and insurance companies seek guidelines as a tool for monitoring patient care for consistent and efficient treatment approach for a specific diagnosis.

1.2.3 Benefits of Clinical Guidelines

1. They improve the quality of patient care by offering explicit recommendations for physicians who are uncertain on how to manage patients with specific diagnosis and they also improve the consistency of care by providing recommendations to physicians who are comfortable in following them as they have been developed from multiple studies which have been proven to benefit the patients.

2. The guidelines have been developed from studies which were multi-centered, randomized, double-blinded studies which have attained clinical significance with mentioning of specific sub population groups (which include African Americans, Caucasians, Hispanics, Asians) being benefited by various diagnostic and therapeutic approaches and give precise instructions on how to treat a specific disease and thus are helpful to physicians who are comfortable in implementing them on their patients as these guidelines are evidence-based.
Healthcare systems benefit from the clinical guidelines in implementing effective management and improving efficiency by standardizing care. Healthcare systems reduce the length of stay of patients in the hospital by following these guidelines and thus decrease the total expenses. Healthcare systems are being paid by insurance companies based on the suggestions made by clinical guidelines for the length of the stay for a specific disease for which a patient is being hospitalized.

- For example, a patient with pneumonia admitted to the hospital should be treated and discharged from the hospital within 4 days according to the guidelines. The insurance companies follow these guidelines and pay the hospital for only four days of stay. The hospitals follow these guidelines and try to discharge the patient within the time frame thus reducing the total expenses.

1.2.4 Developing Clinical Guidelines

Development of clinical guidelines is based on the research evidence obtained from various studies conducted in various centers throughout the country and also incorporates the views of experts who are mostly physicians in the particular specialty. The guidelines are developed from the subjective assessments of the research evidence based on a subset of patient population.

In the development of a clinical guideline, the first step would be identifying and refining the subject area of the guideline. The topic for the guideline development usually is by a dialogue amongst physicians, patients and the potential users of the guidelines like the insurance companies.
2 Second step would be convening and managing the guideline development groups.

3 On the basis of systematic reviews the group assesses evidence about the clinical question or condition. This usually involves reviewing of various studies conducted and clinical trials performed. The evidence is then translated into a recommendation within a clinical practice guideline.

4 The last step in the development of a guideline is external review of the guideline. Guidelines usually should be reviewed by external reviewers who are people with expertise in clinical content who can review the guideline to ensure clinical sensibility and experts in systematic reviews who can review the method by which the guideline was developed.

1.2.5 Computerized guidelines

Based on the guidelines for management of specific diseases, software has been developed which includes various options as a checklist to be filled in by the physician who deems them as appropriate for a patient. These options are usually developed from various guidelines for the specific disease and the physician has to go through the checklist and select the appropriate treatment options. By doing so the physician goes through the most commonly suggested guidelines for a specific disease and has to make a clinical decision on whether to start the patient on a particular management option and hence the possibility of missing out a guideline is low.

1.3 Problem Statement
Can we figure out if a particular guideline has been followed or not from looking at the discharge summaries? Natural language processing techniques are used to process clinical records to identify patients with congestive heart failure and to evaluate them against clinical guidelines suggested by Milliman Care Guidelines. The results are then being cross checked with a clinician’s expertise evaluation of the electronic medical records.

1.4 Organization

In Chapter 2, the different ways of representing guidelines and the various natural language processing applications in healthcare are discussed. In Chapter 3, a natural language processing based approach to check the clinical records against the guidelines represented in a particular format is proposed and the results are discussed in chapter 4. Finally in chapter 5, the pitfalls in it and the future work in the field is presented.
Chapter 2 Literature Survey

2.1 Clinical Practice Guideline Representation Models

2.1.1 Introduction

The demand for reducing variation in medical practice and increasing economic pressures has increased attention on representing clinical practice guidelines electronically. Since all the published guidelines cannot be memorized by a physician, an automated method for delivering the guideline’s logic along with access to clinical and patient data will be very helpful.

Representing clinical guidelines in a format that can be shared across applications and organizations is a key challenge [1]. Research is being carried out in inventing various approaches to model the guidelines. These approaches have to be designed in such a way that they have to provide optimum performance (by including access to structured clinical and patient data) in clinical practice and should also provide interoperability within a number of information systems operating under the NHII (National Health Information Infrastructure). Due to a lack of standardized way of representing guidelines and ambiguity in text-based guidelines, few guidelines can be implemented in clinical information systems.

A good guideline representation model is the one which is unambiguous and makes use of stored patient data. A number of tools and computational models for extracting, organizing, presenting and sharing clinical guidelines are currently in use.

2.1.2 Developing Computer Interpretable Guidelines from Text
Representations of guidelines written on paper or on computer are referred to as external representation and the image of a guideline in a person’s mind as internal representation. The representation of a guideline must be informationally and computationally equivalent to the textual guideline [3] i.e. they should contain the same concepts and relations and should lead to same cognitive steps and operations. The representation should also allow flexibility so that practitioners who are geographically distributed can access the information.

Vimla L.Patel et al. [3] analyzed the various theories and methods of cognitive science used in generating clinical guidelines. A physician would represent a textual guideline to get an internal representation by using his clinical knowledge and experience. An expert focuses only on the necessary steps and high level concepts. On the other hand, computer scientists come up with a representation that limits itself to medical concepts that are obvious from the given textual guideline. An expert cannot represent the guideline which can be understood by everyone and an external representation (a literal transformation of text) given by a non expert does not include the logic and alternatives that are not inherent from the text. For the representation to be accurate and efficient, the guideline representation team that includes both clinicians and experts in computer-based representations is preferable.

2.1.3 Requirements of a Guideline Model

(1) Making decisions, (2) sequencing actions and decisions, (3) defining goals (e.g., achieve/maintain/avoid special patient states), (4) interpreting data and (5) refining actions are the five tasks identified by Tu and Musen that should be provided by a computer-supported guideline based care. According to Shiffman et al. [4] no matter for
what purpose, with what goals and for whom (users or applications) the guidelines are developed; the basic requirements of the model are that it should be comprehensive, expressively adequate, flexible, shareable and reusable.

Elements in a guideline representation model are called primitives. Dongwen Wang et al [7], after reviewing various guideline representation models, identified that most of them included decisions and actions as primitives in their representation. A recommendation in a clinical task or intervention is defined as an action. Based on some pre-defined criteria in a guideline, selection from a set of alternatives leads to a decision. Patient states and execution states were the other popular primitives widely used. Patient state describes the state of a treated individual based on the actions performed and the decisions made within the context of a guideline. Based on the stages of process with regard to decisions and actions, the guideline implementation system’s execution state can be described. Some of them even provide primitives for information collection.

Representation of guideline structures is also important. Most of the representation models represent the guidelines as temporal constraints and nesting of guidelines. Nesting of guidelines is done by capturing the composition of a complex guideline into sub guidelines. Temporal order of primitives is given by the temporal constraints for e.g., whether they have to be executed sequentially, in parallel or in iterative manner. Clinical guidelines can be put to best practice if they are given access to patient data. So a provision for access to patient data should also be provided in the model.

2.1.4 Guideline Representation Models and tools
2.1.4.1 GLIF: Guideline Interchange Format (GLIF) is a specification for structured representation of guidelines developed by the InterMed Collaboratory. [8] GLIF consists of a number of object class definitions each carrying some data. A guideline class looks like the following.

```
Interface Guideline {
    attribute string name;
    attribute sequence <string> authors;
    attribute string intention;
    attribute Criterion eligibility_criteria;
    attribute sequence <Guideline_Step> steps;
    attribute Guideline_Step first_step;
    attribute sequence <Supplemental_Material> didactics;
};
```

The guideline traversal is started from a starting point and later traversal is based on the values obtained from the action or decision steps. The intention of a guideline is free text which describes the purpose of the guideline. Eligibility criteria are a list of conditions and didactics consisting of pointers to supporting materials for the guideline.

The guideline step can be a decision step or an action step and each of them is separately defined as a class with its own attributes like sub guidelines and patient data for action step and condition steps, branch steps and sequence of steps for a decision making step. Synchronization steps are used along with branch steps in order to achieve flow of control through various parallel paths.

2.1.4.2 Arden Syntax

The Arden syntax for Medical Logic Modules (MLMs) is a standard adopted by American Society for Testing and Materials (ASTM) to encode medical knowledge. [9] Sufficient logic is provided in each MLM to make a single medical decision and they run
automatically using an appropriate computer program called event monitor. The structure of the syntax has slots in each of the three categories, maintenance, library and knowledge. The maintenance category consists of information about the document, author etc…, the purpose of the MLM is represented in a textual manner in the library section and the knowledge category contains slots for logic to embed if-then rules, action slot etc. No specific slot for patient data is provided and also it is not convenient to represent complex guidelines that have to be further broken down into sub guidelines.

2.1.4.3 Asbru

Asbru is part of the Asgaard project in which task specific problem solving methods are developed that perform several tasks. It is text-based and is designed as a set of protocols and a guideline is represented as a set of hierarchical skeletal plans. The elements of Asbru are plan-body (which specifies the dynamics of a plan, state transitions like ready, rejected, activated, suspended etc…), conditions of seven types (activate mode, suspend condition, complete condition etc…) and also the order of the conditions (sequential, parallel, unordered and any-order), intentions, preferences, effects, time annotations, arguments and domain definitions. All of these primitives are represented in an xml format.

2.1.4.4 EON

Protocol based care can be represented by EON architecture developed by the Stanford Medical Informatics. It is also called the Dharma model that has access to patient data, domain ontology, medical concepts and guidelines. It structures a guideline using abstract definitions, guideline algorithm, decision models and recommended actions. The EON guidelines are encoded in Protégé 2000. [10] EON is a template based language that
provides form based templates so that common decision criteria can be easily written by domain experts. An inference to generalization/specialization relationship is reached using taxonomic hierarchies provided by the templates and PAL (Protégé axiom language). Appropriate criteria-evaluation engine is invoked by the guideline execution engine for every criterion. The criteria evaluation engine returns the result which consists of truth value of the criterion and annotations.

2.1.4.5 GEM

Guideline Element Model (GEM) is an XML based guideline document model that includes a broad set of concepts that are present in the entire guideline life cycle. [11] The heterogeneous information contained in guidelines is adequately expressed by GEM using its elements which are more than a 100. Major concepts in the model include purpose, intended audience, method of development, knowledge components (recommendations, definitions and algorithms), and testing and review plan which in turn contain sub elements like the recommendations (that contain conditional step, logic step, action step etc….). Using XML for its representation makes it shareable, reusable, flexible and can easily be processed by computers for future use.

2.2 Natural Language Processing (NLP)

2.2.1 Introduction

Human beings can understand natural language easily but this is not the case with computers. Natural language has a lot of expressiveness, great ambiguity, and vagueness. The main problems with language are (1) a single sentence can be expressed in many ways. (2) The meaning of a word depends on the context it is used in. (3) Ambiguity in expressing relations. (4) Some expressions lack specificity. To provide accurate and
reliable data NLP is used. Lot of knowledge is required to understand natural language. People possess significant knowledge and use it to understand language [12].

Natural Language processing applications developed so far have been used primarily for extraction and encoding, in healthcare. NLP applications, in general, include natural language generation, summarization, knowledge representation, knowledge acquisition, search and indexing (Google is a classic example), domain models, machine translation, and grammar and spell checking. NLP has significant number of applications in healthcare as well.

2.2.2 NLP in Healthcare

2.2.2.1 Introduction

Availability of Computerized patient records or Electronic healthcare records (EHRs) enables improved quality of care and reduces costs [13]. Most of the automated applications developed so far access clinical information that is coded. Though numeric data like laboratory data, easily coded pharmacy medications data are readily represented in coded format, not all the information present in clinical records can be coded because most of the clinical information is from discharge summaries, emergency department’s chief complaint and several other reports. All these reports contain information that is free text. This free text cannot be accessed by the automated applications.

The information is locked up within the free text. It is difficult to access the text as the meaning differs based on the context [12]. For example, if we are to search if a medication named X has been given, we cannot simply search for the word X. This is because, the word X can be surrounded by other terms like the medication X has been
discontinued, the patient is no more on X now, he was on X till 1984 etc…. So checking for the keyword along with modifiers that imply the context is also important. In such cases, we can either get the coded data directly (which is a tedious job) or we can use Natural Language Processing techniques to get the required data.

2.2.2.2 Medical Language Processing (MLP)

Application systems that extract and codify information from patient reports are called Medical Language Processing (MLP) applications. Natural language understanding concerns with a variety of components of knowledge namely syntactic, semantic, and domain knowledge components. [14]Although general language resolution is mysterious, Medical Language Processing (MLP) is a well developed field of research because the domain is restricted and constitutes a sublanguage. The sublanguage grammar removes the ambiguity, vagueness, amount of variety as it is restricted to a specific well-defined domain having specific relations and information of objects. The output from the MLP systems is stored in different ways based on its application. Some systems use conceptual graph, which is a graph like form, some use frames which are like database tables. For web-based applications it is a best option to represent the information in Extensible Markup Language (XML).

2.2.2.3 NLP systems developed in Healthcare

2.2.2.3.1 Medical Language Extraction and Encoding System (MedLEE)

The goal of MLP system is to extract, encode and structure the clinical data and translate the information to terms in a controlled library like UMLS or SNOMED, so that it can be used by other automated applications.[18][19] MedLEE uses Medical Entities Dictionary (MED) for synonym resolution and disambiguation of terms. MedLEE is composed of functionally variant modules. Each module encodes and transforms the text in agreement with a particular aspect of language. The four components are preprocessor, parser,
phrase regularizer and encoder. Output from each module is processed by subsequent module until the final structured data is obtained.

2.2.2.3.2 MENELAS

This is a consortium aimed at providing a better access to patient discharge summaries. This NLP pilot system is an access system for medical records. [20] It comprises a morphological, syntactic, semantic and pragmatic analysis of the medical sublanguage. It has an analysis system that analyses the data in medical reports and stores as a normalized, conceptual structure representing its meaning in a database which are then used by target services. [21] MENELAS provides access to the data through services like electronic activity board used by staff; retrieve set of patients with specific characteristics for case based reasoning, test of research hypothesis, and automatic production of nomenclature codes.

2.2.2.3.3 DAD system

It is a discovery tool developed for test based discovery in biomedicine. It is a concept-based NLP system for Pub Med citations. The root for its sprout in the NLP world was a simple but an important concept. Concepts in one domain may be useful in another domain without a mere knowledge of it to the researchers. Those concepts are contained in a condensed form in scientific publications. So a literature-based survey helps the researcher to explore a new domain. The resource for the researcher is a literature database MEDLINE, which gives public access through interfaces like PubMed. The DAD-system architecture is a server-client model. A user throws a query to the system and the system derives a hypothesis using a vast set of scientific citations relevant to the
query. The system is not autonomous but should be initiated by the user. The resources used by the system are PubMed, MetaMap, UMLS knowledge sources.

Lot of research is being carried out in data mining for healthcare, electronic patient record order entry systems, automatic detection of adverse events, automatic encoding of clinical documents and in medical surveillance related applications.

2.2.2.4 Natural Language Processing and Clinical Care guidelines

2.2.2.4.1 Automating severity score guideline for community acquired pneumonia employing MLP on discharge summaries

Carol Friedman et al. [17] did a retrospective study on community acquired pneumonia (CAP) on discharge summaries from a hospital in New York and classified patients into five groups based on their risk factors. They developed an application using NLP techniques to extract information from discharge summaries and assign patients automatically to various groups based on their risk factors to help physicians identify low-risk patients and thus determine duration of hospital stay. The study involved an independent physician’s help to evaluate the accuracy of the application by manually verifying the discharge summaries. The accuracy, sensitivity and specificity were found to be 93%, 92%, 93% respectively. This study demonstrated that by using NLP techniques, it is feasible to automate determination of risk classes for patients with CAP.

2.2.2.4.2 NLP in the EMR: assessing clinician adherence to tobacco treatment guidelines

In a retrospective study conducted by Hazlehurst et al. [24] to assess clinician adherence to tobacco treatment guidelines by using MediClass (a NLP tool for MLP). They found that the MediClass performance was sufficient to assess the clinician’s adherence when compared to human coders who were manually interpreting the charts to
see if the guidelines have been followed for the treatment. They developed a protocol from the guidelines to evaluate the clinician’s adherence to them.

2.2.2.4.3 Extracting information on pneumonia in infants using natural language processing of radiology reports

Eneida A. Mendonca et al. [25] conducted a study by developing a clinical application using NLP techniques in detection of Healthcare associated pneumonia in new born which demonstrated that automated method was feasible in comparison to clinician’s judgement with a sensitivity of 87.5% and a specificity of 94.1%.

2.2.2.4.4 Electronic Medical Records for Clinical Research: Application to the Identification of Heart Failure

In a study conducted by Serguei Pakhomov et al. [26] to identify patients with heart failure through EMR by using NLP techniques they found out that NLP approach had a sensitivity of 100%. They searched for the terms congestive heart failure, checked if ejection fraction < 40%, heart failure, biventricular failure, cardiomyopathy and other related terms. The resulting documents were filtered by checking if these words were surrounded by a negative context in a range of +/- 7 words. This study proved that NLP is a more suitable and faster approach in identifying patients with heart failure when compared to that of predictive modeling approach.
Chapter 3 Design Approach

3.1 Design

3.1.1 The Framework

The framework architecture is shown in Figure 1.

3.1.1.1 Research Corpus

The corpus was obtained from WVU Hospitals through a standard IRB process. 1000 discharge summaries were used for this study. The 1000 discharge summaries were de-identified using De-Id software (Gupta et al. 2004). A discharge summary is a textual
clinical record. It contains information about admission diagnosis, discharge diagnosis, discharge medications, active medical problems at discharge and various procedures and complications in the hospital. A sample discharge summary is shown in Figure 2. This study is restricted to discharge summaries related to Congestive Heart Failure (CHF).

Figure 2. A sample discharge summary

3.1.1.2 Congestive Heart Failure

Congestive heart failure (CHF) is a leading cause for admission into the hospital with over one million hospitalizations per year [23] and also is a significant cause for mortality
in the United States. It affects an estimated 2 million Americans and is responsible for approximately 20,000 deaths per year [23]. The treatment of CHF represents a major portion of expenses in the health care system. The impact of CHF on health and health care expenditures urged for the development of guidelines.

The vast number of studies done so far on congestive heart failure suggest that patients who are being admitted with this problem into the hospital should be evaluated for left ventricular systolic function by an echo cardiogram during hospitalization or planned for an echo cardiogram in the near future after the discharge and patients who have left ventricular dysfunction should be started on Angiotensin converting enzymes (ACE) inhibitors or Angiotensin receptor blockers until and unless they are contraindicated.

A significant decrease in mortality and morbidity was observed when the above discussed guidelines were followed. Clinical guidelines recommending evaluation of echo for CHF exacerbation and starting the patient on ACE inhibitors evolved. Various quality measures have also been implemented by the healthcare systems based on these guidelines for monitoring the patient care. But still some patients who are being admitted with congestive heart failure are neither being evaluated for left ventricular dysfunction nor being started on ACE inhibitors. The percentage of heart failure patients having evaluation of left ventricular ejection fraction in 2004 was 81.6% and those prescribed ACE inhibitors at discharge were only 64.8% (National Healthcare Quality Report, 2006).

This clearly indicates a need to monitor whether the quality measures have been implemented on these patients at the time of discharge. At present this is being done by
manually checking the discharge summaries of various patients who are being admitted for congestive heart failure. An automated way to validate the discharge summaries against the clinical care guidelines for would be of great value to the Healthcare system.

### 3.1.1.3 Clinical Guidelines for CHF

Milliman Care Guidelines has licensing agreements with more than 1000 clients across the country to provide its clinical guidelines through a variety of software options. [27] The care of one in three Americans is managed using Milliman Care Guidelines. The work presented here is based on their 10\(^{th}\) edition of Inpatient and Surgical care guidelines for CHF. The focus of this research has been on quality measures (Figure 3).

The Milliman care guidelines for congestive heart failure (CHF) suggest that patients who are admitted with diagnosis of CHF should have an echo cardiogram to assess left ventricular function before arrival, during hospitalization or at least planned for after discharge. All the patients with left ventricular systolic dysfunction should be on ACE inhibitor (Angiotensin converting enzyme) and/or Angiotensin receptor blocker at hospital discharge until and unless they are contra indicated. The contra indications for ACE inhibitors include angioneurotic edema, pregnancy, renal artery stenosis and ACE inhibitor related allergic reactions. An ACE inhibitor is a type of medicine used to treat heart attacks, heart failures or a decreased function on the left side of the heart.
3.1.1.4 Flowchart Representation of the Guidelines

A flowchart representing the above information about guidelines is developed (figure 4).
Figure 4 Flowchart Representation of Clinical Guidelines for CHF
3.1.1.5 XML representation of the Guidelines

Information in the flowchart is represented in an XML format. A snippet of the representation is shown in figure 4. It has tags for filtering criteria (used to filter the documents relevant to CHF), steps of the guidelines and also a special tag for contraindications. The element contents of these tags have the keywords of the guidelines.

```xml
<?xml version="1.0"?>
<clinical_guidelines>
    <congestive_heart_failure>
        <filter_documents_criteria>
            <criteria>Congestive</criteria>
            <criteria>cardiomyopathy</criteria>
            <criteria>CHF</criteria>
        </filter_documents_criteria>
        <guidelines>
            <step1>
                <recommendation>echo</recommendation>
                <recommendation>echocardiogram</recommendation>
                <recommendation>Echocardiogram</recommendation>
                <recommendation>Echo</recommendation>
                <recommendation>Ejection</recommendation>
                <recommendation>ejection</recommendation>
            </step1>
            <step2>
                <recommendation>Ejection</recommendation>
                <recommendation>ejection</recommendation>
            </step2>
        </guidelines>
    </congestive_heart_failure>
</clinical_guidelines>
```

Figure 5 XML representation of the Guidelines

3.1.2 Evaluation

3.1.2.1 Approach for Filtering and Validation
The approach used by the java based framework is to search for keywords in the xml tags, get bigrams of the keyword, check if the keyword is associated with a negative context and check for values in the bigrams if any.

3.1.2.1.1 Bigrams

Bigrams are group of two words. For example strong tea, most powerful etc….Bigrams of a keyword are found within a window size of +/- 10. These bigrams are used in determining the context of the keyword and also values of the keyword (wherever appropriate).

3.1.2.1.2 Context of the keyword

Following Firth’s quote ‘you judge a word by the company it keeps’; the contextual meaning of the word is determined by checking if the keyword is surrounded by negative phrases. Wendy Chapman identified seven phrases that account to 90% of the negations in narrative clinical reports are no, denies, without, not, no evidence, with no and negative for [41]. These phrases are searched for in the bigrams of the keyword. The presence of negative phrases in the bigrams of the keyword implies that the keyword is surrounded by a negative context.

3.1.2.1.3 Values associated with keywords

In order to check for the value of ejection fraction (EF), initially the bigrams of the keywords ‘EF’ or ‘ejection’ are checked for the word ‘normal’. If it is not found, then the presence of a two digit number whose value is less than 40 (this is an indication for congestive heart failure) is searched using a regular expression.

3.1.2.1.4 ACE/ARB
The documents are searched for ACE/ARBs against a list embedded in one of the guideline tabs. The list consists of the various ACE/ARB and their generic names adopted from the U.S. department of Health and Human Services [42, 43]. If the medications are not mentioned, then the document is checked for contraindications to the therapy. The contraindications are specified under a special tag <contraindications>. The output of the system is the sentence in the document which mentions the keywords. These sentences are presented to a reviewing physician or nurse to decide if guidelines have been followed.

3.1.3 Assumptions

It is assumed that the physician who dictated the discharge summaries mentions about the care provided for the patients admitted with CHF. This should include mentioning ejection fraction and if the patient is on ACE inhibitors or not and also the contraindications if any.
Chapter 4 Results

4.1 Process used for analysis

After development of the software system described above, we reviewed 1000 discharge summaries. The results were compared to a gold standard generated by a manual review by a physician of the same 1000 discharge summaries. Sensitivity, specificity, positive predictive value, and negative predictive value were all calculated on the results (figure 5).

Figure 6. Process used for analysis

[44]The sensitivity of a test is the probability that it will produce a true positive result when used on an infected population (as compared to a reference or a gold standard). The
specificity of a test is the probability that a test will produce a true negative result when used on a non infected population (as determined by a reference or gold standard). The positive predictive value of a test is the probability that a person is infected when a positive test result is observed. The negative predictive value of a test is the probability that a person is not infected when a negative test result is observed. These values are shown in table 1.

Table 1 Result analysis

<table>
<thead>
<tr>
<th>Test results</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>True positive</td>
<td>False positive</td>
</tr>
<tr>
<td>negative</td>
<td>False negative</td>
<td>True negative</td>
</tr>
<tr>
<td></td>
<td>tp/tp+fn=sensitivity</td>
<td>tn/tn+fp=specificity</td>
</tr>
</tbody>
</table>

4.2 Score

The manual review process found a diagnosis or findings of congestive heart failure in 42 of the 1000 discharge summaries. The values are shown in table 2. The automated process identified 41 of these cases, for a sensitivity of 98%. In addition, the system falsely flagged 3 other charts as having congestive heart failure, giving a positive
predictive value (precision) of 93%. The results obtained from the automatic evaluation by the framework and those from the manual evaluation and the relationship among the terms is shown in figure 5. The automated process identified a vast majority of the cases.

Table 2. Results

<table>
<thead>
<tr>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Test results</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test results</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>True positive = 41</td>
<td>False positive = 3</td>
</tr>
<tr>
<td>negative</td>
<td>False negative = 1</td>
<td>True negative = 955</td>
</tr>
<tr>
<td></td>
<td>Sensitivity = 0.9761</td>
<td>Specificity = 0.9968</td>
</tr>
</tbody>
</table>
4.3 Discussion

The first step of the framework is to identify documents that are relevant to CHF by searching for keywords ‘CHF’, ‘congestive’, ‘cardiomyopathy’ and also makes sure that these words are not surrounded by a negative context.

The framework identified 44 discharge summaries relevant to CHF. However 3 out of the 44 documents were not relevant. In the following discussion, the number shown is the number of the document which runs between 1 and 1000 followed by the context in the document which was misinterpreted by the framework.

False Positive #1: “The patient’s electrolyte and lab values accompanied by a negative chest x-ray, both ruled out possible underlying cause being congestive heart failure”.

Figure 6: Evaluation Results
False Positive #2: “Upon admission here, our initial impression was that she had preeclampsia with no evidence of peripartal cardiomyopathy”.

False Positive #3: “These evaluations were thought to be related to his asthma rather than a cardiomyopathy”.

We have found out that by adding the word ‘ruled out’ and ‘no’ in the negation phrases list during filtering, False Positives 1 and 2 are identified. Identifying False Positive #3 will require a higher level of natural language processing.

The framework was not able to identify one document which was relevant to CHF.

False Negative #1: “On cardiac catheterization, EF was shown to be 35-40%”.

This suggests that the patient had congestive heart failure but there were no keywords which the framework was searching in the documents. Another keyword ‘EF’ when added to filtering criteria, False Negative #1 is identified by the framework.

The evaluation results for the 41 documents identified matched exactly with the gold standard. The framework results suggest that 5 of the discharge summaries did not have any information regarding the prescription of ACE/ARB s at the time of discharge (guideline step3) even though no contraindications were apparent. These discharge summaries were presented to the physician and the manual evaluation suggested that 3 out of the 5 documents did not have information about prescription of the ACE/ARB provided in the list.
Chapter 5

Conclusion

This study demonstrates that the care described in dictated clinical records can be compared to published national guidelines using Natural Language Processing techniques embedded in an XML framework. Use of this technique could help physicians and hospitals monitor quality and optimize patient outcomes. The guidelines can be transformed to a collection of keywords. The study can be extended to multiple guidelines. The XML framework is also extensible. Additional XML tags can be added (wherever needed) and XML attribute values can be used to mention values. The framework after review can be fine tuned with additional keywords according to the feedback.
References


5. S.W.Tu and M.A.Musen. Representation formalisms and computational methods for modeling guideline-based patient care.


7. Dongwen Wang, Mor Peleg, Samson W.Tu, Edward H.Shortlife, Robert A.Greenes Representation of Clinical Practice Guidelines for Computer-Based Implementations


14. Carol Friedman, George Hripcsak, Irina Shablinsky *An Evaluation of Natural Language Processing Methodologies.*


17. Carol Friedman, Lyundmila Shagina, Yves Lussier, George Hripcsak Automating a Severity Score Guideline for Community-acquired Pneumonia Employing Medical Language Processing of Discharge Summaries.


22. Text-Based Discovery in Biomedicine: The Architecture of the DAD-system Marc Weeber, MA, Henny Klein, PhD, Alan R. Aronson, PhDy, James G. Mork, MSy, Lolkje T.W. de Jong - van den Berg, PhD, Rein Vos, MD, PhD.


25. Eneida A. Mendonca, Janet Haas, Lyudmila Shagina, Elaine Larson, Carol Friedman Extracting information on pneumonia in infants using natural language processing of radiology reports


35. Paul Shekelle, Martin P. Eccles, Jeremy M. Grimshaw, Steven H. Woolf. *When should clinical guidelines be updated?*

36. [http://evidence.ahc.umn.edu/health_care_guidelines.htm](http://evidence.ahc.umn.edu/health_care_guidelines.htm)


41. Wendy W. Chapman PhD, Will Bridewell BS, Paul Hanbury BS, Gregory F. Cooper MD PhD and Bruce G. Buchanan PhD. 2002. *Evaluation of Negation Phrases in Narrative Clinical Reports.* Center for Biomedical Informatics,
Computer Science/ Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA.

42. http://www.globalrph.com/ace.htm

43. http://www.medicare.gov/AssistancePrograms/include/dataset/costSavings/TherEquiv/incDrugList.asp

44. http://www.rapid-diagnostics.org/accuracy.htm