Cognitive Computing for Electronic Medical Records

2 authors:

Murthy V Devarakonda
Arizona State University
33 PUBLICATIONS   **452** CITATIONS

Neil Mehta
Cleveland Clinic
51 PUBLICATIONS   **990** CITATIONS

Some of the authors of this publication are also working on these related projects:

Information Management View project

NER in Biomedical Domain View project

Chapter 32
Cognitive Computing for Electronic Medical Records

Murthy V. Devarakonda and Neil Mehta

Abstract The explosive growth of data has led to a situation where the human brain is overloaded with more information than it can process. It is particularly dire in healthcare where critical information may be buried in the mountains of data in the Electronic Medical Record systems (EMR systems) and healthcare workers struggle to make sense of this information to provide the best care for their patients. Cognitive computing, exemplified by Watson, offers the promise of transforming EMR systems from mere data storage to intelligent systems that help physicians in providing improved patient care. When seeing a patient, a physician needs to quickly grasp the summary of the patient's medical history from the EMR to prepare for the visit and to put the patient's complaints in context. During the visit, there may be a need to supplement, confirm, and investigate the information that the patient provides with information from the EMR. These information needs can be fulfilled by a cognitive system using advanced analytics on the patient record data. Some of the ways this can happen are a problem-oriented summary of a patient record, precisely answering natural language questions about the patient record content, automatically identifying urgent abnormalities, and by providing precise causes for such abnormalities. In this cognitive computing view, an EMR is an active entity that leverages the vast knowledge of the medical sciences, drug information, and medical ontologies in the context of the patient medical records to meet the information needs of the healthcare provider.

M.V. Devarakonda, PhD (IBM) (✉)
Research Staff Member and Manager, IBM T.J. Watson Research Center,
Yorktown Heights, NY, USA
e-mail: mdev@us.ibm.com

N. Mehta, MBBS, MS
Education Institute and Medicine Institute, Cleveland Clinic,
Cleveland Clinic Lerner College of Medicine of Case Western Reserve University,
Cleveland, OH, USA

32.1 Introduction

The potential for health information technology to support clinical care and transform the health care delivery system has long been recognized [13]. The HITEC Health Act of 2009 and Meaningful Use incentives starting from 2013 have encouraged integration of health information technology in the clinical setting. While some benefits due to the technology have been observed with the introduction of Electronic Medical Record systems (EMR systems),[1] physicians continue to struggle with potential workflow disruptions and the resulting decrease in productivity in using EMR systems [2]. A recent American Medical Association study has identified reducing cognitive load as one of the priorities for improving usability of electronic health records [1]. This presents a clear need and an opportunity to use advanced analytics, such as those demonstrated by IBM Watson, to improve physician's efficiency and effectiveness in using EMRs.

Expert systems have been developed for medical applications in the past. However, very few have been adapted for practical use, and even fewer have been designed to improve the use of EMR. For example, MYCIN [3] is one of the first research attempts in 1970s to apply artificial intelligence to identify bacterial infections and recommend antibiotics. While it was a successful experiment, it was never used in practice. Isabel [14] is a modern symptom checker system which identifies likely diagnoses from symptoms described in natural language. It does not provide other features of a cognitive computing system mentioned earlier like a semantic search or a problem-orientated medical summary. Recent research work on IBM Watson [8] adapted the system to the medical domain and showed that it can answer medical questions, such as the American College of Physicians' Doctor's Dilemma questions and United States Medical Licensing Examination Step 1 questions, with a high degree of accuracy.

So, why can't the existing expert systems address the cognitive load on physicians in patient encounters? The missing piece of the puzzle is the integration with the EMR data. In all the medical diagnostic expert systems, a user is expected to extract relevant data from an EMR and present it to the system as an input, and there lies a major challenge. This takes precious additional time and effort and it is not easy to determine exactly what information to include. Patients don't have just one medical problem. Providing input relevant to one potential disease may lead to a solution for that one disease but not a holistic solution for patient care. A system that hopes to reduce a physician's cognitive load must be applied to where the key information exists, i.e. the patient's EMR.

In this chapter, we present an approach to applying cognitive computing to EMRs using IBM Watson. We demonstrate the value and feasibility of the approach with an application of IBM Watson called Watson EMRA (Electronic Medical

---

[1] Computer stored and managed patient data is referred to by multiple names, such as, EHR and EMR, often with little or no difference between the terms. To avoid possible confusion, we consistently use the term EMR to refer to a patient medical record and EMRs as its plural. Furthermore, we use the term *EMR system(s)* to refer to the software and hardware system that stores and provides access to the contents of EMRs.

Record Analysis). We begin by providing a background on the concepts of cognitive computing; and then summarize a physician's cognitive needs in an outpatient setting based on interviews with physicians at two major hospital systems. We next discuss a model of patient record summarization based on an automated problem list generation and the use of semantic search within an EMR. The discussion includes a user perspective of the impact of these capabilities alleviating cognitive load. We explore the full potential of cognitive computing for enhancing the use of EMRs and conclude the chapter with a brief summary.

32.2 Cognitive Computing

According to IBM, cognitive computing marks a new era of computing where computers interact with users in a natural way, learn continuously, and expand human cognition [15]. Cognitive computing systems are built from techniques developed over past several decades in many areas of computer science research including, natural language processing, information retrieval, knowledge representation, machine learning, and advanced data analytics. It is a confluence that is enabled by continuous development in computing hardware, software engineering, and many decades of research in algorithms for natural language processing and machine learning. The resulting cognitive computing systems can analyze, predict, reason, and interact with humans in ways that are natural to us. These cognitive computing systems do not aim to eliminate humans in the decision process but instead attempt to augment human intelligence and cognition.

IBM Watson [7], by winning the Jeopardy championship [24], has become *an opening to an era of cognitive computing* according to Kelly and Hamm [15]. Since the Jeopardy event, research has continued at IBM to adapt Watson to the medical domain [8] and to solve realistic problems in patient care. This effort created a powerful foundation, using which new applications can be built to address the cognitive needs of physicians in patient care. Before discussing these applications, let us first explore the cognitive needs of a physician in the next section.

32.3 Physicians' Cognitive Needs

Physicians for the most part follow a typical *workflow* in a patient contact. The concept of workflow was first introduced by Frederick Taylor and Henry Gantt, in late nineteenth century, to bring scientific principles to management of manufacturing [23]. This work gave raise to time and motion studies which became a systematic methodology to optimize manufacturing processes and service delivery. While specific details of a clinical workflow distinctly varies from specialty to specialty and from one individual physician to another even in the same specialty, there are certain high level steps that are consistently repeated in a typical patient contact.

A closer examination of these workflow steps help us identify a physician's cognitive needs and how solutions to these cognitive needs impact the overall outcome of physician efficiency and patient care. Since patient care takes place in many settings, in order to arrive at a practical solution let us examine a physician's workflow in one common setting, i.e. in outpatient care. Shartzer divides clinical tasks performed by a physician [19] into four distinct steps:

1. Visit Preparation
2. Patient History and Examination
3. Assessment and Plan
4. Visit Wrap-up

These four steps form a general workflow for an outpatient clinical setting. These tasks along with the transitory steps are the key to understanding physicians' information needs. An ongoing study [17], developed from interviews with physicians of two major hospital systems from a broad range of specialties, further breaks down the information needs at each step.

The first step, the visit preparation, involves a review of patient profile, problems list, event notifications, and routine activities. Here a physician is seeking information such as: "What was done at the last visit? What data has accumulated since the last visit? What is overdue or needs to be addressed today?" It is necessary for a physician to be able to find important information without being overwhelmed with irrelevant information. At this stage an *abstracted* summary would be useful as it would avoid physician having to read through several previous notes, lab results, procedures, and medication orders to formulate the abstraction in their own minds.

In the second (history and physical) step, a physician's information needs can be broadly described as filling gaps in the patient's history (*supplementation*), verification of something either reported by the patient, stated in the medical record, or suspected by the physician (*confirmation*), and exploring how or why a diagnosis or treatment evolved (*investigation*)..

In the "Assessment and plan" step, as the physician is evaluating test results, coming up with a diagnosis, contemplating further tests, and developing a treatment plan, she may rely on various sources of knowledge. Of course, one of the key sources is their own medical knowledge, but they may also want to look up information that could be relevant, such as current guidelines, newly developed treatment options, and ongoing clinical trials. The information available to the physician in this situation should be highly focused on the specific issue at hand and contextually related to the specific patient, and not a general document or web page with broad (and possibly irrelevant) information. Lack of precision and relevance in the available information at this stage leads to distraction, irritation, and inefficiency instead of intelligent assistance.

At the end of the visit, physicians want to make sure that everything that is needed to be addressed has been addressed. This includes not only the reason for the visit or the chief complaint which is typically addressed with a treatment or management plan in the earlier step, but also any routine activities or outstanding health maintenance items. In addition, physicians strive to provide a clear decision and

communication to the patient on what the next steps are. It is also the time to document or at least prepare for documentation of the patient visit at a later time. In this stage of the workflow, the cognitive needs are highly focused around what steps the patient needs to take (such as medications and preparation for diagnostics if any) and could significantly impact the outcome from proposed plans and documentation thereof.

While this workflow describes only one patient contact scenario, i.e. an outpatient visit, it is a concrete example of a physician's information needs, and therefore, a prime target for cognitive computing solutions. One such solution, an automatically generated problem-oriented patient record summary [6] is described in the next section. It is intended to help physicians in the patient visit workflow by providing a quick summary of a patient record and the ability to browse for specific information.

## 32.4 Problem-Oriented Patient Record Summary

As Weed [21] pointed out several decades ago, a medical record should be organized by a patient's medical problems for it to be useful in their diagnosis, treatment, and management. He called it a problem-oriented medical record. Given the centrality of the medical problems, it would be natural and effective to model patient record summary around them. But, the problem list is rarely well maintained and so physicians find it usually unreliable. A reliable problem list is needed as a part of patient record summary, and we will discuss an automatic extraction of a problem list using natural language processing and machine learning later in the chapter.

Previous approaches to clinical summarization involved applying a succession of aggregation, organization, reduction/transformation, interpretation, and synthesis to a specific patient data. Linear abstraction works well for a lab result or a single patient problem, but a summary of an entire EMR needs to go beyond this. For example, it also needs to inter-relate such individual data as we discuss below. So, the natural way to achieve coverage while maintaining brevity is to start with aggregates of key patient data types such as problems, medications, labs, encounter notes, and procedures, and then provide additional semantics over them.

EMRA summarization therefore consists of multiple clinical aggregates, including the problem list, medications, clinical encounters, and lab results. Elements of each of these may be aggregated to some level by themselves. For example, results of a lab may be organized, transformed and interpreted such that the summary shows the latest value and an indication as to whether it is now, or has ever been, out of the normal range.

As mentioned above, there are also important relationships between the data aggregates and need to be surfaced. For example, a problem is treated by one or more medications. Neither the problem data aggregate nor the medications data aggregate reliably contains such clinical associations. These relationships may not be explicitly documented in a visit note either, even though they are the result of a

Fig. 32.1 The Watson patient record summary model showing the generated problems list and other clinical data aggregates along with clinical relationships among them
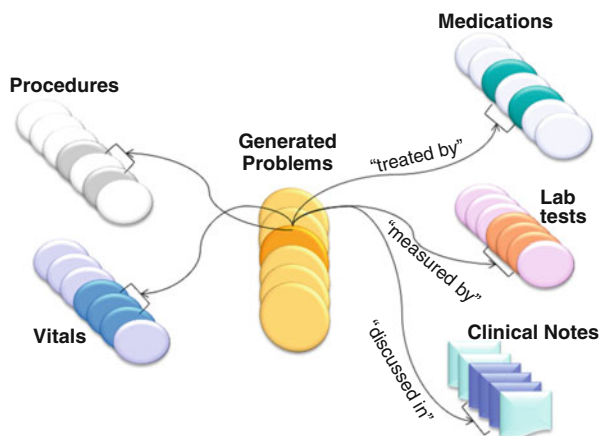
physician's judgment and actions. So, identifying such relationships between the problem list and other clinical aggregate are a part of the summarization.

Do elements of a clinical aggregate have an association? For example, two of a patient's medications may be closely related by the fact that they both treat the same problem. There are also other aspects such as the pharmacologic mechanisms of action of a medication and pharmacologic effects on human physiology, and so intra-relationships among medications are complex, and similarly two problems on the problem list may be related in multiple ways. In general, however, some elements of a list may have a closer relationship with each other than with the others. Physicians make these associations instinctively and based on their training. An intelligent summary of a patient record should present data aggregates in a clinically meaningful manner. EMRA summarization produces a nearness score based on multiple intrinsic relations among elements of an aggregate which identifies how closely an element is related to the other elements of the aggregate. For example, this analysis allows us to present the medications list in a clinically meaningful manner.

Encounter notes are a unique data aggregate in an EMR. They are the notes written by physicians, nurses, and other clinicians on every contact with the patient. Some of them may simply capture notes of a telephone call with the patient, and the others may involve detailed notes of a physician in a comprehensive visit. From the information content view point, not only is the data within a clinical note valuable, but the existence of the note and the data describing it (known as *meta-data*) are also equally valuable. The existence of notes in a time period indicates the amount of care provided. The meta-data may include the specialty and the note's author and the type of the note (i.e. whether it is a Progress Note or a Discharge Summary) further expanding on what of type of care received and who provided it. Therefore, our summarization organizes the clinical notes by specialty and by timeline, identifies the note type, and relates them to the problem list. Watson EMR analytics (specifically, the automated problem list generation algorithm) produces the association between each of the problems listed and one or more clinical notes. Figure 32.1 shows an abstract model of our summarization comprehensively

Fig. 32.2 A dashboard-style visualization of the Watson patient record summary, showing clinical data in tables and patient contacts as a timeline

representing the analysis and the organization of clinical data discussed so far. A Web-based Graphical User Interface view, implementing the summarization model, is shown in Fig. 32.2 for an actual patient record.

32.5 Using Patient Record Summary for Patient Care

This patient record summary can meet important cognitive needs identified earlier in the chapter. Let us consider an Internal Medicine physician seeing a patient with diabetes.. The physician may not have seen the patient before or it may have been several months since the last visit. The physician needs to learn or recall the patient's medical history somewhat quickly prior to the visit. The Watson EMRA patient record summarization helps the physician in visit preparation by presenting an accurate and reliable problem list, along with the active medications, labs, vitals, and recent visits to physicians and hospitals. In preparing for this patient visit, physician notices that the patient has Diabetes Mellitus Type II along with comorbidities Dyslipidemias, Obesity, and Microalbuminuria from the summary view (Fig. 32.2). Noticing related comorbidities is made easier because Watson EMRA shows them close to Diabetes in the problem list.

Next, the physician observes clinical associations of the Diabetes with other clinical data of the patient by clicking the checkbox next to Diabetes in the problem list. As shown in Fig. 32.3, upon selecting the problem, related patient's medications – Metformin and Glipizide, in this case – are highlighted and shown at the top of the list. Figure 32.4 shows an isolated view of the problem – medications association. In addition, related labs and clinical encounters are highlighted. The highlighted lab results show the most recent value and indicate if the value is within the normal range

Fig. 32.3 When a medical problem is selected, the dashboard highlights related patient data including medications, labs, patient contacts, and procedures
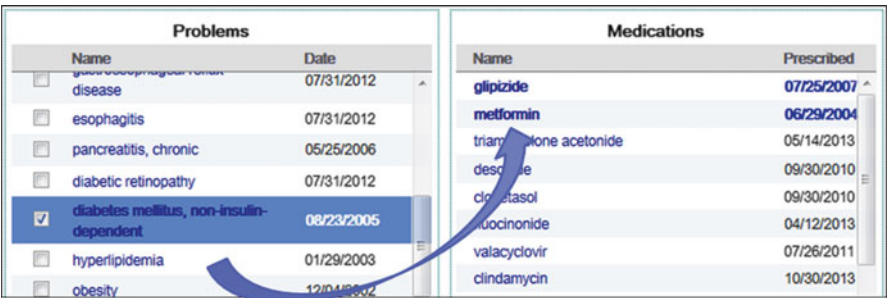
Fig. 32.4 The problem and medication relationships are isolated here for clarity, and note that the related medications are highlighted and moved to the top of the list

as per the ranges defined in the lab test panel. Viewing this summary provides a rapid understanding of the patient's treatments and labs for the problem and relevant notes from previous encounters. We should note that the problem to clinical data associations are not in the EMR but are generated by Watson EMRA using novel analytics based on natural language processing techniques adapted for the medical domain.

In the history and physical step, the physician needing help with supplementation, confirmation, and investigation can find the necessary clinical data details either directly in the summary view or by detailed data by at most two clicks. For instance, let's say the physician would like to investigate historic glycemic control as indicated by Hemoglobin A1c over time, he/she can click on Hemoglobin A1c in the labs table, which opens a new window showing the historical values of the lab (see Fig. 32.5).

Now the physician wants to confirm what was planned by the primary care physician or Internal Medicine specialist in the most recent visit related to Diabetes, the

Fig. 32.5 From the summary dashboard, one click enables access to detailed lab test results, for example, here Hemoglobin A1C data is shown as a plot over time along with reference high and low values, and as a table

Fig. 32.6 Access to physician notes is also available with one click from the summary dashboard, and the selected note is shown with relevant problems highlighted

physician goes to the clinical encounters table, finds the encounter categorized under primary care and highlight as related to Diabetes, and clicks the marker which opens a window showing the clinical note (See Fig. 32.6). The note provides the

information needed the physician is looking for. The same note can also be accessed by clicking on the problem (Diabetes) from the problem list. A list of relevant clinical notes appears, each with a brief synopsis, and the physician can preview the synopsis and then click to fully open the needed clinical note. In each clinical note, Watson EMRA highlights references to the problem and so reading the note for details on Diabetes history, observations, assessment and plan is made easier. This association between a problem and clinical notes is also enabled by the Watson EMRA analytics.

In the assessment and plan stage, the physician needs highly focused information as they decide on a course of action. Let's say, the physician in this case is thinking of introducing an additional medication to improve A1c and blood sugar levels. He/she might want to see if the patient was on the medication before. The medications table allows switching to discontinued medications so that the physician can see if the medication was given and discontinued before. If the physician wants to know why it was discontinued he/she can use another function Watson EMRA called Semantic Find, which will be described later in the chapter. In addition, the physician wants to ensure the patient is prescribed medication for hypothyroidism (which he notices as comorbidity from the problem list) and that TSH levels are at desirable levels, which he/she do by clicking the box next to hypothyroidism in the problem list.

In the final visit wrap up, the summary view provides the physician with the necessary context to write the new encounter note. This context includes the problem list, active medications, and current labs. If necessary, he/she can review previous notes selected by the specialty and timeline. Later in the chapter we will discuss a semantic search functionality of Watson EMRA, which will also be useful in these visit workflow steps. In the next section, we will take a deeper look at how Watson EMRA generates the problem list [5] and how physicians can use this understanding in making the best use of the generated problem list.

32.6   Automatic Problem List Generation

Most EMR systems allow physicians and clinicians to enter and maintain the problem list manually. However, this problem list is not usually well maintained and as a result physicians almost always ignore it [4, 11, 12, 18]. There are many reasons for this state of affairs which include lack of proper support from the EMR systems, lack of clarity of what goes on the list and what comes off of the list, multiple authors populating the list, and many intended uses of the list, at least some of which are contradictory. Perhaps the fundamental reason, which is often missed in the discussions of the EMR systems, is that the problem list maintenance is a knowledge and time intensive task requiring significant investment of an experts' time. If for the sake of argument we set aside the difficulty of the problem list creation and maintenance, it is indisputable that the potential value of an accurate problem list is considerable.
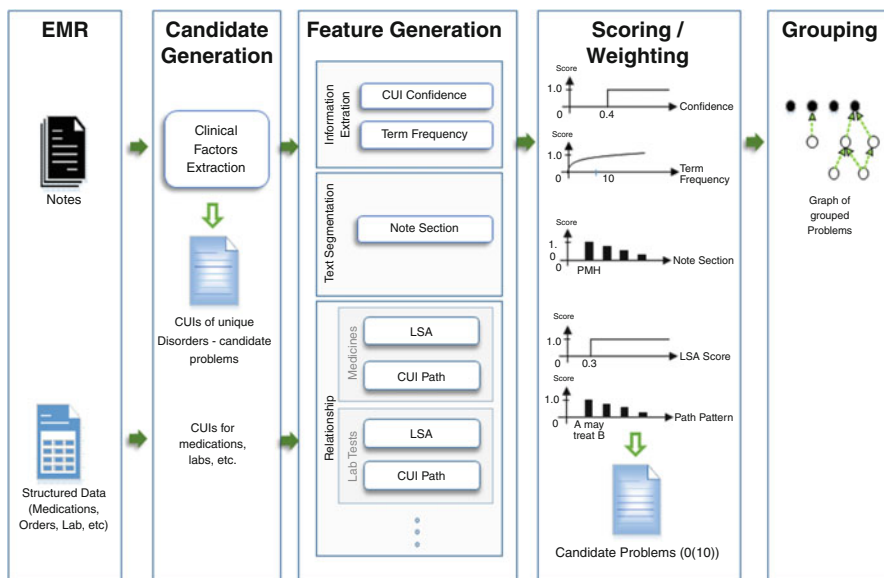
Fig. 32.7 The Watson problem list generation uses natural language processing to extract features from the patient record that are used with a machine learning model to generate the problem list

The EMRA problem list generation starts with an automated step of identifying a large pool of medical disorders mentioned in the encounter notes of a patient's EMR. It then goes through additional steps of algorithmically gathering evidence for each potential problem, and then in the final two steps the candidate list is reduced to a final and presumably an accurate problem list and closely related problems are merged (See Fig. 32.7). The EMRA method uses NLP and machine learning. These steps are described in some detail below.

Watson EMRA recognizes the words and phrases denoting medical disorders in the encounter notes and assigns one or more Concept Unique Identifiers (CUIs) from the UMLS Metathesaurus [20]. This internal representation of words and phrases allows reasoning about them as medical concepts, such as recognizing medical synonyms, i.e. recognizing that HTN, high blood pressure, and hypertension all represent the same disease. In fact, Watson EMRA recognizes all medical terms in the clinical text, not just disorders, and categorizes them into UMLS semantic groups, e.g. as Disorders, Chemicals and Drugs, and Procedures. Each of these groups is further subcategorized, for example, Disorders are sub-grouped as Diseases or Syndromes, Signs or Symptoms, Findings, and others. Mapping terms (i.e. words or phrases) to CUIs is, in itself an interesting research task because the mapping between terms and CUIs is many-to-many, and the correct CUI may depend on the context. So, in addition to using the standard natural language processing methods and UMLS lookup, Watson takes advantages of additional context and sentence structure to obtain better mapping, and uses a numerical score to

indicate the confidence that a CUI represents a given term. This confidence score is one of many features used in the problem list generation as discussed below.

In the first step of the method Watson EMRA identifies a term in an encounter note as a candidate problem if the term is categorized in the above CUI mapping process as a diseases or a syndrome, or one of a select set of findings. For a typical EMR, this results in identifying a few hundred candidate problems. When compared to the final list, the list of candidate problems has high recall (>90 %) but poor precision (<10 %). We note that recall represents the percentage of correct problems reported and precision represents the percentage of reported problems that are correct. So, this initial step attempts to capture all the correct problems but it may also include many problems that are not correct. The subsequent steps attempt to improve precision of the problem list without a substantial loss of recall.

In the next step, the method produces a set of feature values which will be used in a machine learning model in the next step. Longitudinal EMRs are a rich source of information and extracting and aggregating the information into the features is crucial to success. We used many types of features – lexical, medical, frequency, structural, and temporal features – each which we will describe below.

32.6.1 Lexical Features

We used the standard TF-IDF (term frequency – inverse document frequency) formulation, where the term frequency is number of occurrences of a term (candidate problem) normalized using the maximum frequency of any term in the document and the inverse document frequency is the inverse of the fraction of documents with the term in the corpus. Depending on the goal, a document can be a note or an EMR. When generating the problem list for a patient, an EMR is a document and the entire collection of EMRs is our corpus. When deciding which encounter note is relevant to a selected problem, the encounter note becomes the document and an EMR becomes the corpus. For the problem list generation, IDF is calculated using the entire de-identified EMR collection.

Unlike a normal text document, an EMR is a longitudinal record and therefore, more recent notes are likely to better represent the patient's medical problems. Also, each note in the EMR has implicit sections and so a term (e.g. hypertension) appearing in different sections (e.g. family history vs. assessment and plan) may have significantly different meanings. Because of this, in addition to calculating TF at the EMR level, TF is also calculated for each note section and for a few different time periods.

32.6.2 Medical Features

Terms in the EMR semi-structured data are also mapped to UMLS CUIs so that we can use the UMLS relations. Medications turn out to be one of the most important features, whereas lab tests and procedure orders were less useful. One reason is that

the medication names are relatively standardized, even while mixing the generic and brand names, and a UMLS CUI can be reliably found. Conversely, labs and procedures are often specified in institution specific abbreviations instead of CPT codes and LOINC codes, and are therefore harder to accurately map to UMLS concepts. Another reason is that while medications are prescribed to treat problems, some lab tests are very general and the others are very specific. For example, Hemoglobin A1c is used only to check for blood sugar control while a Basic Metabolic Panel could be ordered for glucose, calcium, potassium, renal function, and others. The relation between a problem and a medication is derived from a weighted confidence score obtained from distributional semantics and UMLS relationships.

32.6.3 Problem Frequency Features

Certain problems occur commonly among a patient population, and thus the frequency of a problem can be thought of as the prior probability that the patient is likely to have it. Two sources of frequency are used in our method. The first is the SNOMED CORE usage, which represents the frequency in a broad population. The second is calculated using all diagnosed problems (as ICD-9 codes) in our collection of EMRs, which represents the frequency in a particular institution.

32.6.4 Structural Features

The concept "diabetes mellitus" appearing in the assessment and plan (informal) section in a patient's progress note is a much stronger indicator that the patient has the disorder than the same concept detected in the family history section in a nursing note. Since notes are in plain text and note metadata is optional, the structures have to be learned. Watson EMRA detects informal sections in a note using regular expressions and heuristics, and the informal section in which a term appears is used as a feature.

32.6.5 Temporal Features

The span of an EMR varies from a single day to several decades. Most temporal features in our experiments are normalized to prevent bias towards longer EMRs, but the absolute value is also used to define certain features, e.g. note *recency*, where the recency is defined as the number of days from the most recent patient contact.

Temporal data elements are used in three ways. First, they are used as features directly. Temporal features considered include the first/last mention of a problem, and the duration of a problem. Second, the temporal data is used to align semi-structured data and unstructured data, e.g. a medication prescribed before a problem is mentioned in a note is not considered as evidence to the problem. Third, temporal data is

used to divide notes into bins on the timeline so that frequency can be counted by intervals, e.g. term frequency in recent notes vs. term frequency in earlier notes.

32.6.6 Machine Learning Model

Once all feature values are generated, they are converted to numerical values and normalized to a standard 0–1 scale. Subsequently a machine learning algorithm, the Alternating Decision Tree [9] generates a confidence score for each potential problem, and problems with a confidence score above a threshold are accepted as the entries on the patient's problem list. Both the machine learning model and the confidence threshold are *learned* using a gold standard we developed with the help of medical experts.

32.6.7 Gold Standard

To evaluate the accuracy of the Watson problem list generation method, we tasked medical experts to create a gold standard using initially 199 EMRs (which later grew to 400 EMRs) acquired from the Cleveland Clinic under an IRB protocol for the study. The medical experts, mainly medical students in the fourth year of their medical degree program, studied the EMR, including the encounter notes, medications ordered, labs, procedures, and allergies, created a problem list. Each EMR was reviewed by at least two medical students and they separately created two problem lists. Next a physician has reviewed the lists and adjudicated any differences between the two lists.

The final gold standard still needs further refinement to be useful in training and testing our method. The problem lists created by the medical experts are usually in English terms that require mapping to UMLS concept unique ids or CUIs. We decided to use SNOMED CT CORE (US National Library of Medicine 2014) as the vocabulary for the problem list as this vocabulary has been developed for the express purpose of being used for the problem list. Therefore, we needed to map the gold standard developed by the medical experts to the SNOMED CT CORE, and usually this mapping required further review because of the many-to-many mapping between textual problem terms and the SNOMED CT subset. We set aside a test set of 20 % of random EMRs from the gold standard and used them to assess the accuracy of the Watson method.

32.6.8 Candidate Problems

Figure 32.8 shows a distribution of the number of candidate problems generated per EMR (across all EMRs in our test and train set). We see a nearly normal distribution, with an average of 135 candidate problems and a standard deviation of 33. The

Fig. 32.8 The candidate problems per patient record in the data have a near normal distribution with a median of about 140 candidate problems per patient
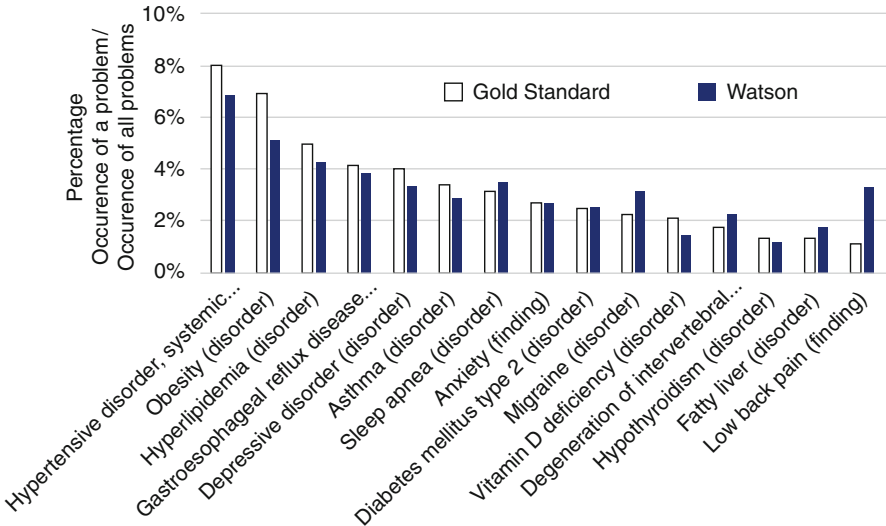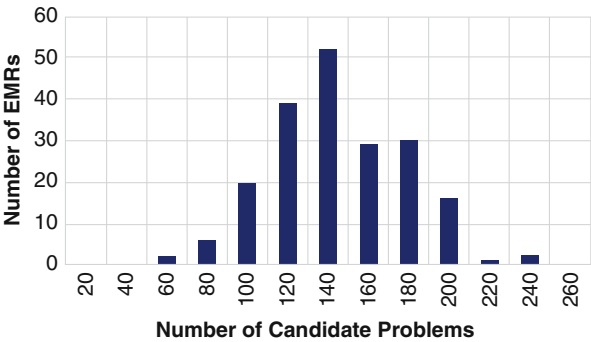
Fig. 32.9 The Watson problem list generation accurately identifies most of the top 15 frequent problems occurring in the gold standard as seen in the bar chart below

machine learning model reduces these candidate problems to an average of 9 predicted final problems, a reduction by over 93 %.

32.6.9 Most Frequent Problems

Figure 32.9 shows the 15 most frequently occurring problems and their frequency in the gold standard. Juxtaposed against them, Fig. 32.9 also shows how closely Watson EMRA's prediction tracks the gold standard for these most frequent problems. Watson EMRA is mostly accurate in predicting frequently occurring problems. However, our model does not do well with lower back pain. A problem like

Table 32.1 An accuracy analysis of the Watson problem list generation method indicates promising results with a recall (sensitivity) of 80 % when optimized for high recall

Model prediction objective	Recall (%)	Precision (%)	F1 score	F2 score
Tuned for maximum F1 score	70	67	0.69	0.69
Tuned for maximum F2 score	80	53	0.64	0.73

Fig. 32.10 The top two levels of the alternating decision trees machine learning model used in the Watson problem list generation
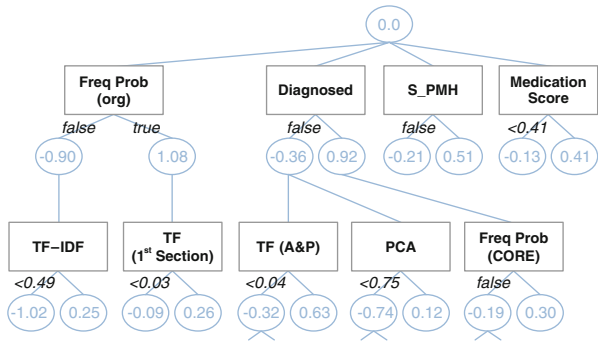
this is usually a challenge for our model. Physicians often prescribe medications for this especially when it is acute or severe, but subsequently if it is chronic a patient may be taking over the counter medications that may not be listed in the medications list or controlling this with back exercises. In including the problem in the gold standard medical experts used somewhat non-specific reasons, such as the severity and there not being another problem that explains the finding. Overall, EMRA accuracy on the most frequent problems is very good.

32.6.10   Overall Accuracy

At this time, the Watson EMRA achieves a recall of 70 % and the precision of 67 % on this gold standard as shown in Table 32.1. What it implies is that on average roughly 70 % of actual problems are captured in the list generated and 67 % of the problem list entries are correct. It is possible to tune the method so that it provides a higher recall and slightly lower precision while keeping the overall "accuracy" same, which ensures more of the actual problems at the risk of introducing more noise in the problem list generated.

32.6.11   Features with the Strongest Contribution

Which machine learning features have the strongest positive contribution for correct predictions in the Watson EMRA method? Figure 32.10 shows the top two levels of the Alternating Decision Tree machine learning model used in the

Watson EMRA method. From the Figure we see that the problem frequency (i.e. how common a problem is), whether it is in the diagnosis codes of the EMR, whether the problem is in the previous medical history of a note (S_PMH), and whether the patient is being treated with a medication for the problem are the features with strongest influence on the model. This observation shows that our model well captures the basis a physician might use in reviewing an EMR to identify the patient's problem list.

32.7 How to Use and Interpret the Generated Problem List?

There will always be a margin of error in a computed result, but with the help of evidence created for a problem selection in the problem list generation process, it is possible to examine the evidence and use human judgment before accepting the results for patient care. A part of the evidence for a problem is the set of clinical notes that mention the problem or its clinical synonym. An examination of the notes would reveal if the problem was identified by a physician or if it was a false positive. In the latter case it, the physician would instruct the system to ignore it and the system would learn from the feedback.

Another part of the evidence is the feature values of the machine learning model for problems. An examination of the weighted feature values typically reveals which features were responsible for a candidate problem to become a problem list item. A closer examination of the dominant feature reveals whether the score was justified or not. If the score seems inappropriately high, a physician can once again provide feedback to the system which will help correct the selection. In spite of the need to verify the results, the generated problem list offers a practical and efficient way to maintain and use the problem list in clinical practice.

32.8 Semantic Search for Clinical Information

Summarization described above may not address all the information needs of the physician. Studies [10] indicate that while browsing is a predominant mode of information seeking, search is often employed when browsing fails to produce the desired result. So, when a physician is looking for specific information that is not provided in the summarization, a search function is needed to fulfill the information need. For example, if a diabetic patient's previous labs indicate microalbuminuria, a physician treating the patient may now want to know if the patient was an ACE inhibitor. If not, why not? In general, this is a level of detail that is not usually available in the summary. But a search of the patient record can provide this information. Manually scanning through the record is not only tedious but also error prone.

Watson EMRA provides a search function that takes a set of words as input and finds matches for the search terms on many *semantic* dimensions. We call this Semantic *Find* to emphasize its similarity to finding matches in a document based on

clinical similarities, not just textual matches. For example, Semantic Find identifies exact ("literal") matches of the search terms just as any standard document search, but even more importantly it also finds clinical synonyms. Searching for "hypertension" would match clinically equivalent terms such as "BP Elevated", "high blood pressure", and even a report of blood pressure measure of 147/95 in the EMR.

Semantic Find also finds other useful types of matches such as more general and more specific matches. If one enters "back pain", it will of course returns instances of semantic matches to "backache" but it also returns instances of "lower back pain" as a more specific match and instances of "pain" as a more general match. These matches help in seeing a broader scope of matches related to the search terms, and may be helpful in determining a new treatment or modifying an existing one.

In medicine, absence of certain findings is almost as important and may be even more important than the presence of the findings. Take for instance, the finding of Deep Vein Thrombosis (DVT) in a patient after a recent surgery. Determining its absence is important for treatment as well as for gathering quality metrics. Semantic Find therefore searches for negated instances (such as "no DVT") when searching for a term (i.e. DVT) and identifies them as negated results in the output.

To provide rapid response, Semantic Find builds an internal index of all medical concepts recognized in an EMR. The index construction is made possible by the Watson analytics that recognize all words and phrases which represent medical concepts in an EMR. When a search is initiated on an EMR, the search terms are also mapped to one or more medical concepts using the Watson analytics and the concepts are "looked up" in the EMR's medical concepts index. Different ways of looking up or comparing the search concepts yields a different facet of the search results. For instance, synonymy comparison of the search concepts with the concepts index yields semantic matches. The hyponym relationship yields more general matches and the hypernym relationship yields more specific matches. This matching takes place in the context of UMLS – i.e. depends on the relationships UMLS defines for a pair of concepts. See Fig. 32.11 for the results of Semantic Find for the search term "back pain" in an EMR, and notice how different tabs provide matches along different dimensions.

Multiple hyponym/hypernym relationships may be defined in UMLS which results in multiple matches along this dimension. Furthermore it is possible to mix synonymy with hyponym/hypernym relationships and find even more indirect but still relevant matches. These complex matches can quickly become expensive and slow the response time, and so we employ heuristics to limit the search. Semantic Find also provides matches for the terms in the semi-structured data in an EMR such as in the Ordered Medications list. The "treats" relationship from UMLS is used when the search term is a disease or a symptom.

32.9 Using Semantic Find to Meet Cognitive Needs

The overall value of the semantic find can be seen in its ability to complement the patient record summary in meeting a physician's information needs in the workflow of a patient contact. While the patient summary provides a quick way to understand
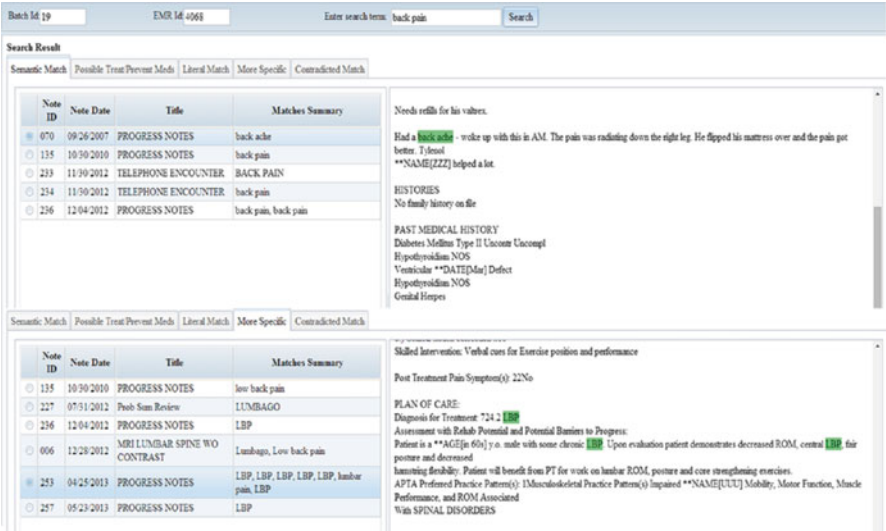
Fig. 32.11 The Watson patient record Semantic Find matches search terms to the contents of a patient record on several dimensions, including medical semantic match, more specific and less specific matches, and contradicted matches

the patient's overall clinical status, Semantic Find helps to probe the record for specific information that may not be available in the summary. This can be particularly important in supplementation, confirmation, and investigation during the history and examination phase.

Semantic Find is also useful in finding specific information during Assessment and Plan. For example, if the physician is looking for an answer to the question – Why did the patient stop taking medication *Sitagliptin?* (After the physician finds that the patient discontinued the medication from the patient summary.) The physician enters the medication name as a search query, and when the results are presented, he/she looks for the most recent (by date) result in the literal or semantic matches returned by the search.

Similarly, to find an answer to the question: Did the patient complain about sleeplessness before December 2013? The physician enters the symptom (in this case *sleeplessness*) and looks at the literal, semantic, more specific, and more general matches up to the specific month and year. The semantic find Graphical User Interface helps this process by displaying results for each type of match in a different tab and by listing results in reverse chronological order.

32.10 Looking into the Future

From these current capabilities we can build new and more sophisticated capabilities in the system to expand the assistance a cognitive system can provide to a physician. These advanced capabilities reduce the amount of work a physician needs to

do in using an EMR for patient care, and the cognitive system takes on an increasing responsibility to provide highly specific and targeted cognitive help.

32.11   Natural Language Question Answering on an EMR

Semantic Find is a powerful capability in finding clinically semantic matches in an EMR for given search term or terms. When the search terms succinctly capture the information needs, it delivers the needed results. However, when the information need can only be specified as a natural language question with all its inherent nuances, an advanced Question Answering capability is needed including the system capability to understand the question correctly and then find the relevant answer(s) precisely. Watson has demonstrated this ability even in the medical domain when the target of the question is the medical knowledge as represented by the text corpus provided to Watson. However, answering questions in a similar way when the target of the question is a single EMR is a distinctly different challenge at a technical level, and is an active area of research at IBM.

32.12   Advanced Patient Summary

The patient record summary presents the problem list for a patient and relates it to other clinical data aggregates, but a physician may need more detailed information about a specific problem in the list. For example, if the problem is the hypertensive disease, the physician may want to know what the duration of the disease was, and if there was any end organ damage such as its manifestations on kidneys or heart. The physician may want to know the timeline of blood pressure readings and if any medications were added, removed, or changed overtime. For some other types of problems like headache, it is important to understand if the problem is recurrent, chronic, or acute. Is there a plan in place, is there a definite diagnosis, or is there a need to monitor and follow up on the problem? From a cognitive computing perspective, these are advanced information extraction and abstraction challenges. Some of the data such as the medication time line is available from the semi-structured data but the majority of the information needs to be identified in the unstructured text, abstracted as necessary, and reasoned about. Watson EMRA is an excellent foundation to build these additional capabilities.

32.13   Guidance on Treatment Options

Weed proposed Knowledge Couplers as a way to improve physician's decision process during patient care. The idea behind the Knowledge Couplers is that they automatically apply rules representing the medical knowledge to patient care workflow steps. At the history and physical step of the visit, this knowledge guides a physician

on what data to collect. The collected data plus additional medical knowledge helps physician decide on a set of diagnostic tests needed. Subsequently, the history and physical data and the assessment from the diagnostic tests as well as additional medical knowledge helps the physicians decide on treatment plans. This powerful conceptual model can be realized using Watson core capabilities of reasoning with medical knowledge [16] and Watson EMRA capabilities of analyzing an EMR. When realized using the Watson technology, this capability can help a physician by prompting "have you considered this?" as they are exploring the next steps in diagnosing and treating a patient condition. It can bring a wealth of latest treatment guidelines, medical knowledge, and specific patient data such as comorbidities, symptoms, and current medications to bear upon the consideration of next steps.

32.14 Guideline Extraction from Documents

In determining the treatment options, how will the medical knowledge become available in a form that can be used in automated reasoning? Some efforts are directed towards a manual process of human experts creating these knowledge rules. While at first it seems a reasonable and expedient way to do so, one realizes very quickly that it is highly human resource intensive, brittle, and difficult to update and correct. After N rules exist in the system, adding a new rule requires understanding and assessing its impact on various combinations of existing N rules. The number of combinations to consider grows very quickly even for small numbers of N, eventually leading to combinatorial explosion that is way beyond human cognition. Therefore, our approach to generating machine usable knowledge from guidelines document is to use natural language processing for extracting the knowledge. There is early work demonstrating the feasibility of this approach [22].

32.15 Summary

In this chapter, we discussed how the principles of cognitive computing are realized in Watson EMRA, a patient record summarization and semantic search capability built on the foundations of Watson. The functionality of the Watson EMRA is driven by the information needs of physicians in patient care. Watson EMRA takes a longitudinal patient record and creates a summary of the record, centered about an automatically generated problem list. The problem list is generated using natural language processing and machine learning. The summary also includes semantic relations between the problems and other clinical data aggregates such as medications ordered. For the times when the patient summary is not adequate for finding specific details, Watson EMRA also provides a semantic search which finds matching semantic medical concepts in the semi-structured and unstructured EMR contents, along several dimensions, including more general, more specific, and negated instances. The future work in this area includes advanced summary of problem

status, natural language question answering on an EMR, and cognitive assistance to a physician in terms of next steps in diagnostic testing and treatment. The technology described here is a proof point of cognitive computing for Electronic Medical Records, and an indication of future promise.

References

1. American Medical Association. Improving care: priorities to improve electronic health record usability. 2014. [Online]. https://download.ama-assn.org/resources/doc/ps2/x-pub/ehr-priorities.pdf. Accessed 9 Oct 2014.
2. Bowens FM, Frye PA, Jones WA, 2010. Health Information Technology: Integration of Clinical Workflow into Meaningful Use of Electronic Health Records, Perspectives in Health Information Management / AHIMA, American Health Information Management Association, 7(Fall), 1d.
3. Buchanan GB, Shortliffe HE. Rule based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Boston: Addison-Wesley Longman Publishing Co; 1984.
4. Campbell JR. Strategies for Problem List Implementation in a Complex Clinical Enterprise. Lake Buena Vista, FL, American Medical Informatics Association (AMIA). 1998.
5. Devarakonda M, Tsou C-H. Automated problem list generation from electronic medical records in IBM Watson. Austin: AAAI (Innovatibe Applications of AI Track); 2015.
6. Devarakonda M, Zhang D, Ching-Huei T, Bornea M. Problem-oriented patient record summary: an early report on a Watson application. Natal: IEEE Healthcom; 2014.
7. Ferrucci D, et al. Building Watson: an overview of the DeepQA project. AI Mag. 2010;31(3): 59–79.
8. Ferrucci D, et al. Watson: beyond jeopardy!. Artificial Intelligence. 2013;199(200):93–105.
9. Freund Y, Mason L. The Alternating Decision Tree Algorithm. San Francisco, Proc. of the 16th Int'l Conf on Machine Learning. 1999.
10. Gibbs WJ. Examining users on news provider web sites: a review of methodology. J Usability Stud. 2008;3(3):129–48.
11. Holmes C. The Problem List beyond Meaningful Use, Part I. Journal of American Health Information Management Association, 2011a;81(2), 30–3.
12. Holmes C. The Problem List beyond Meaningful Use, Part 2. Journal of American Health Information Management Association, 2011b; 81(3), 32–5.
13. Institute of Medicine. Crossing the quality chasm: a New Health System for the 21st century. Washington, DC: The National Academics Press; 2001.
14. Isabel Healthcare. Isabel symptom checker. 2014. [Online]. http://symptomchecker.isabel-healthcare.com/. Accessed 22 Dec 2014.
15. Kelly JE, Hamm S. Smart machines: IBM's Watson and the era of cognitive computing. 2013. s.l.:Columbia Business School Publishing, New York, NY.
16. Lally A, et al. WatsonPaths: scenario-based question answering and inference over unstructured information. Yorktown Heights: IBM Research; 2014 (Report No. RC25489).

17. Liang J, Devarakonda M, Mehta N. Cognitive needs of physicians: a study based on interviews at two major hospitals s.l. 2015. IBM Research Report (Manuscript in Preparation).
18. Meystre SM, Haug PJ. Randomized controlled trial of an automated problem. International Journal of Medical Informatics, 2008;77,602–12.
19. Shartzer LJ. Health IT and workflow in small physicians' practices. Quest Answer Brief. 2005. http://www.nihcm.org/pdf/AHRQ-QandA.pdf [Accessed: 7/21/15]
20. US National Library of Medicine, 2009. UMLS Reference Manual. [Online] Available at: http://www.ncbi.nlm.nih.gov/books/NBK9675/ [Accessed 15 04 2014]
21. Weed LL. Medical records that guide and teach. New Engl J Med, 1968;278(11):593–600.
22. Wenzina R, Kaiser K. Using TimeML to support the modeling of computerized clinical guidelines. In: Lovis C, et al., editors. e-Health – for continuity of care. s.l.:2014 European Federation for Medical Informatics and IOS Press. Washington, DC.
23. Wikipedia – Workflow. Workflow. 2014. [Online]. http://en.wikipedia.org/wiki/Workflow. Accessed 24 Dec 2014.
24. Wikipedia. Jeopardy!. 2014. [Online]. http://en.wikipedia.org/wiki/Jeopardy!. Accessed 14 Dec 2014.