Extracting Structured Medication Event Information from Discharge Summaries

Sigfried Gold, MFA, MA, Noémie Elhadad, PhD, Xinxin Zhu, MD, MS,
James J. Cimino, MD, George Hripcsak, MD, MS
Columbia University, Department of Biomedical Informatics, New York, NY

Abstract

We present a method that extracts medication information from discharge summaries. The program relies on parsing rules written as a set of regular expressions and on a user-configurable drug lexicon. Our evaluation shows a precision of 94% and recall of 83% in the extraction of medication information. We use a broader definition of medication information than previous studies, including drug names appearing with and without dosage information, misspelled drug names, and contextual information.

Introduction

We present a parser that extracts structured medication event information from discharge summaries. Uses for such a tool are numerous: quality improvements in clinical information,[1] pharmacogenetics and pharmacogenetics research,[2] pharmaco-epidemiology,[3] decision support and clinical support,[4] pharmacovigilance and post-market surveillance; medication reconciliation; and as an input or pre-processing step for more general natural language processing (NLP) or medical language processing (MLP) tools like MedLEE.[5]

In some medical record systems (c.f., Sirohi[2] and Shah,[3]) medication information will only be available in narrative format and some form of information extraction will be necessary to obtain any medication information at all. However, even in settings with a computerized physician order entry system (CPOE), there may be medication information available in narrative form that does not appear in the CPOE, particularly medications reported on admission and prescriptions given on discharge. Informal inspection of our institution's electronic medical records has shown a considerable difference between what appears in the CPOE and what appears in the narrative record.

We focus on discharge summaries as challenging, characteristic examples of clinical text. Discharge summaries at NYPH normally consist of admission/discharge diagnoses, chief complaint, history of present illness, allergies, past medical or surgical history, medications, social history, physical examination, laboratory data, and hospital course.

Related Work

Only a few published papers address the issue of extracting medication information from narrative clinical text. Evans[1] showed that drug and dosage information could be extracted from clinical notes using the CLARIT™ NLP system. We use their study as a model and compare our method to theirs. However, their system was a proof of concept and did not attempt to extract medication information that would be needed by a production application.

Shah[3] discusses an algorithm that extracts daily dose information from semi-structured text. Their algorithm relies on the medication information already being divided into fields for drug name, quantity, duration, and dosage instructions. They parse the dosage instructions into structured fields, but they do not attempt to recognize drugs and dosages from narrative text.

Sirohi[2] demonstrates the necessity of careful lexicon selection in NLP extraction of drug information. This finding informed our parser design.

There are also commercial systems that extract dosage information (by which we mean drug name, quantity, route, frequency and necessity) from medical records including: LifeCode™, A-Life Medical, Inc. (San Diego, CA); Natural Language Patient Record™, Dictaphone Corporation (Stratford, CT); and FreePharma™, Language and Computing, NV, (Sint-Denjis-Westrem, Belgium). The algorithms of these systems are not generally available to the public.

Our parser was built as part of a larger project, the Medication Extraction and Reconciliation Knowledge Instrument (MERKI). The goal of MERKI is to extract and manage structured and narrative data from a clinical data warehouse; identify drug names as synonymous; perform temporal reasoning on drug information to determine when a patient was on a drug; and remove redundancy so that a single course of a single drug, which may be mentioned in CPOE orders and clinical notes many times, will be condensed into a single data item. Overall, MERKI provides a range of presentation and data export options to allow use in

research, quality assurance, and medication reconciliation settings.

Methods

We generally followed the methodology of Evans.[1] We: 1) defined the concepts to be extracted; 2) built the parser and parsing rules; 3) prepared data for testing; 4) had two physicians annotate the test data to create a gold standard; 5) processed test data with the parser; 6) scored the results. The parser has the following features:

- As a task-specific parser, its algorithm and parsing rules can be restricted to a discrete domain;
- It takes as parameter a lexicon of drug names—in our design of experiments we rely on a lexicon derived from RxNorm entries in the UMLS;
- It takes as parameter a set of parsing rules formulated as regular expressions, which can be tailored for specific tasks or clinical note styles;
- By looking for surrounding dosage language, the parser can find misspelled drug names or drugs that are not in the lexicon;
- The parser can extract relevant information in addition to the drug name:
  - whether the drug was mentioned as an allergy or as discontinued or as prescribed and taken;
  - where or when the drug was administered—at home, in the emergency room, preoperatively;
  - dose, route, frequency and necessity;
  - meta-data from the drug lexicon (e.g., RxNorm CUI).

Concept Definition

Our goal is to provide accurate, comprehensive information about what drugs a patient has been on based on evidence appearing in textual notes. We generally follow Evans, et al. in conceptualizing drug information—a drug expression can consist of a drug name accompanied by dose level, route, frequency, or necessity—however we loosen the concept so that we can catch important medication events that Evans, et al. ignored. Our goal includes identifying:

- drug expressions with misspelled drug names ("Azithromyin 250 mg PO");
- drug names appearing independent of other drug expression elements ("Patient was given K-Dur");
- references to medication events that use drug classes rather than drug names ("Outpt course po Abx", "Pt will remain on antiseizure medications");
- clear implication of medication events ("The patient was subsequently digitalized.")

In order to avoid reporting a drug event where none occurred, such as the mention of a drug the patient refused, or drugs appearing in an allergy list, Evans required a drug name to be accompanied by dose-level, frequency or another one of the attributes they deemed sufficient. We chose to relax this requirement in order to avoid missing true medication events.

We address the problem of drug name appearance where no drug event occurred by defining a context in which the drug name appears. We also use context to help determine the dates of drug administration. In capturing context information we cast a wide net and collect any language we can identify that informs the context (examples from the gold standard and the parser include: "at home", "cur med", "increased", "held", "1980s", "since 13", "hospital course, OR"). We then map this variety of contextual information into four classes: history, hospital course, on discharge, and not administered.

Figure 1 shows a sample of discharge summary text and our breakdown of the medication event into specific fields.

> The patient was discharged on Zosyn 2 grams IV q8 times 12 days and follow-up appointment in the Pediatric Surgery Clinic in two weeks.
>
> Context: discharge
> Drug name: Zosyn
> Dose: 2 grams
> Route: IV
> Frequency: q8 times 12 days
> Necessity: N/A

Figure 1. Sample text from discharge summary with interpretation of a medication event

Parsing Algorithm

Unlike many types of text faced by NLP or MLP applications, medication event information is generally not recursive; it consists of a few types of term—medication name, dose, route, frequency, etc.—strung together in a somewhat inconsistent order, but it does not contain clauses and sub-clauses: it constitutes a regular language rather than a context-free language. We modeled this language through manual analysis, and took advantage of its relative simplicity by constructing parsing rules out of regular expressions as described below.

Following our operational definition of drugs, the parser extracts three types of objects: drugs, possible drugs, and context clues. We use a drug lexicon derived from RxNorm entries in the UMLS to identify drug names. A drug is defined as a drug name optionally accompanied by dosage information (dose, route, frequency, necessity). A possible drug is defined as any non-drug-name text surrounded by enough dosage information to indicate that it refers to a misspelled drug name or a drug name not in our lexicon. Context clues are pieces of text that give us an idea of where, when, and whether the drug was administered

Drug name identification uses a multi-word lookup from the input text to the drug lexicon. The drug lexicon may contain meta-data about the drug name, which will then be connected to the drug in the parser output. As our lexicon is derived from UMLS, we attach the UMLS Concept Unique Identifier (CUI) to drug names. This is not important to the parsing process per se, but the calling program can make use of this information when data is returned regarding the drugs that have been extracted.

Parsing of a clinical note begins by identifying drug names using the lexicon. The rest of the process uses parsing rules stored in a grammar file. The grammar file allows for the definition of terminals and non-terminals. A terminal definition associates a symbol with one or more regular expressions. For instance, the terminal *unitOfMeasure* is defined (partly) as "(g|gm|mg|milligram|mcg|microgram)". Non-terminals are defined essentially in the same way as terminals except that their regular expressions can include symbols for terminals and other non-terminals, which will be expanded into appropriate regular expressions and interpolated into the regular expression of the non-terminal. For instance, part of the definition of the non-terminal *dose* is "number\s*unitOfMeasure?\b\s*form", which will match a number, optional whitespace, an optional unitOfMeasure, a word boundary, more optional whitespace, and finally a form.

The parser returns drugs, possible drugs and context clues in a first pass. It then passes over the returned drug and possible drug text to break these down into their constituent parts.

The context clues extracted by the parser can be supplemented by information from outside the parsed text. In our electronic records system this was necessary because discharge summaries are stored in pieces with separate, named sections, and sometimes the section name gives a better context clue than the text of the section. (Examples of section names are "Attending physician", "Medications", "History of present illness".) Context clues parsed from the text and derived separately are combined, normalized, and attached to the drugs.

Data Preparation

We created the test data set from 26 de-identified discharge summaries from patients seen in 2004 at Columbia University Medical Center/New York Presbyterian Hospital. The test data were left unseen before scoring the results. The development of the parser was carried out using a separate set of unannotated discharge summaries.

Gold Standard Annotation

Two physicians created the gold standard by sequentially annotating the test data: one physician annotated the data and her results were given to a second physician who checked and revised her work. Revisions were discussed with and accepted by the first physician.

Annotation was performed using evaluation forms containing a discharge summary and a table with columns for Drug Name, Context, Dose, Route, Frequency and Necessity. The physicians were instructed to fill in these columns with values quoted verbatim from the discharge summary.

Scoring Parsing Results

The 26 discharge summaries were processed by the parser and drug names were scored as being: 1) a match, 2) a false positive, or 3) a false negative.

The other fields—context, dose, route, frequency and necessity—were scored as one of: 1) Correct, 2) Partially Correct, and 3) Wrong. "Partially correct" was used in instances where the parser produced some but not all of the information provided by the gold standard, for example, if the gold standard recorded a dose as beginning at 5mg and decreasing to 2.5 mg while the parser only recorded 5 mg. Since the context results were catagorical, they were scored simply as Correct ot Incorrect.

Results

The physicians creating the gold standard identified 252 medication events in 26 discharge summaries. For each of these they recorded the drug name or class, context and, insofar as they had evidence for additional attributes, dose, route, frequency and necessity. Forty-seven percent of the medication events in the gold standard had no accompanying dose, route, frequency or necessity.

Out of 252 medication events in the gold standard, the parser accurately matched 208, missed 44 (false negatives), and identified 13 events erroneously

(false positives). This gives a precision of 94.1% (95% CI .90 to .97) and a recall of 82.5% (95% CI .77 to .87).

For the 208 events correctly identified by the parser, the remaining attributes were also compared with the results shown in Table 1.

	Context	Dose	Route	Frequency	Necessity
Correct	65.9%	83.7%	88.0%	83.2%	98.6%
Partially Correct	N/A	5.8%	0.0%	1.4%	1.0%
Wrong	34.1%	10.6%	12.0%	15.4%	0.5%

Table 1. Attribute results

Drug class	27	61.4%
Not in lexicon	12	27.3%
Different spelling	2	4.5%
Software bug	2	4.5%
Misspelled	1	2.3%

Table 2. Breakdown of false negatives

In lexicon	9	69.2%
Software bug	4	30.8%

Table 3. Breakdown of false positives

Tables 2 and 3 show a breakdown of false negative and false positive results by reason.

Discussion

Our precision and recall are comparable to the results reported by Evans,[1] however, as described above, our drug definition is broader than theirs. The Evans drug definition would have identified 96 drug expressions (medication events) in our 26 discharge summaries, whereas our gold standard identified 252.

Shah[3] reports 98.3% accuracy in determining dose and frequency given a valid dosage string. This result, however is not comparable to ours given the semi-structured nature of their data. Our input, in contrast, is unstructured, narrative text.

Drug Name Identification

The largest category of drug identification failure was for drug classes (e.g., antibiotics, antihypertensives, opiods, multi-vitamins.) We made a point of including these in our gold standard because they are important medication events that the parser should catch. However, as mentioned above, parser development up to this point in the evaluation occurred without the benefit of annotated discharge summaries, so the parsing rule file simply does not have rules for catching these. Now that we know what at least some of them look like, we will modify the parsing rules to catch them.

The second largest class of false negatives is caused by drugs not appearing in the lexicon (e.g., cannabis, 10% dextrose, Cardiolite, crack, nebutol). As per Sirohi and Peissig's demonstration of the necessity of careful lexicon selection in NLP extraction of drug information we made the lexicon a parameter of our system.

We are making the MERKI parser publicly available (see Conclusion) along with the same drug lexicon that we used in the evaluation. Although a commercial lexicon or a lexicon based on copyrighted parts of the UMLS might have produced better results, we wanted to use a lexicon that we could make public along with the rest of the parser.

False negatives resulting from different spelling (e.g., the gold standard had Klorcon, the lexicon had Klor-con) can also be addressed by lexicon improvement.

False negatives like the one resulting from misspelling (Labetolol) can be addressed by improvements in the rules for recognizing possible drugs or by introducing fuzzy matching.

Further examination is required to discover why the parser failed to recognize an instance of Naprosyn and an instance of Percocet, both of which are in the lexicon.

The false positives that occurred because they appeared in the lexicon were: air, AT 10, cholesterol, Enterococcus faecalis, oxygen, and sodium. This type of error is due to the sometimes conflicting definition of a medication, and varies from one lexicon to the other.

Four records in the parser results were counted as false positives because invalid text was extracted for the drug name, however, these four records were caught by the matching rules for possible drugs, and these cases were actually valid medication events. We will attempt to improve the parser's ability to extract appropriate text for cases like these, but even where the parser fails to extract a reasonable drug name, these records serve to point a human reviewer to a place in the clinical note where a medication event occurred.

Attribute Recognition

Seventy-one out of 208 (34%) of the contexts were identified incorrectly. Context identification is the most difficult thing the parser attempts to do. Further research is needed to determine whether performance in this area can be significantly improved or if this

task is better left to a general MLP parser like MedLEE.

The primary reason for dose being parsed wrong was that it appeared in the text surrounded by parentheses, which the parsing rules did not account for.

The only reason for a partial score on dose was that doses appeared in the text with more than one value (indicating that the dose should be increased, decreased or alternated); the parser only caught one of the multiple values. The same issue appeared with frequency.

Errors in route occurred for two main reasons: 1) the parser interpreted the word "as" as a route (meaning left ear); 2) the medication expression was formulated in a way that the parser rules did not anticipate.

Like most of the other attribute problems, problems with frequency and necessity were generally caused by patterns appearing in the medication expression that had not been seen and anticipated by the software developer. The parsing rules will be refined to fix these and similar cases over time.

Limitations

The parser was developed without the benefit of annotated discharge summaries. The annotation performed for our evaluation brought to our attention the need for several changes to the parsing rules. Further clinical note annotation will reveal more patterns for the parser to catch.

The MERKI parser extracts medication event information from narrative clinical records, but is only of limited use by itself in determining the drugs a patient actually used. To accomplish that task, the parser output should be reconciled with CPOE and Medication Administration Record data through drug name alignment and temporal reasoning.

The parser was only tested on discharge summaries at New York Presbyterian Hospital; however, we believe that these documents are typical of discharge summaries at other institutions. Further, we believe that our methods will be applicable to other types of notes that are typically simpler in structure and lexicon . The evaluation was based on a gold standard created by cooperation between two reviewers, rather than on one based on agreement between two (or more) independent raters; we cannot, therefore, measure interrater reliability of the gold standard. Nevertheless, the gold standard does reflect at least a valid expert opinion of the medication phrases in the text.

Conclusion

A primary concern in building MERKI has been to create a set of tools that can be used by researchers at Columbia University and at other institutions, so that extracting medication events from electronic medical records becomes a solved problem rather than an expensive and time-consuming tangent to other research.

Our drug event definition builds on previous work, but is more flexible to account for the many different ways in which medication can be referred to in a clinical note. Our parser relies on a library of regular expressions and a lexicon of drug names to identify medication information. Both the lexicon and the parsing rules are flexible, and can be easily customized for other types of clinical notes, or other discharge summaries with different writing styles.

Parser source code is available on the dbmi.columbia.edu domain. Search for "merki medication parser".

Acknowledgments

We would like to acknowledge the contributions of Carol Friedman, Herb Chase and grants R01 LM006910, R01 LM007659 and R01 LM008635 from the National Library of Medicine. Dr. Cimino was supported in part by funds from the intramural research program at the NIH Clinical Center and the National Library of Medicine (NLM). Sigfried Gold is now at Phase Forward/Lincoln Technologies.

References

1. Evans DA, Brownlow ND, Hersh WR, Campbell EM. Automating Concept Identification in the Electronic Medical Record: An Experiment in Extracting Dosage Information. AMIA 1996 Symposium Proceedings. 1996:388-92.
2. Sirohi E, Peissig P. Study of Effect of Drug Lexicons on Medication Extraction from Electronic Medical Records. Pacific Symposium on Biocomputing. 2005;10:308-18.
3. Shah AD, Martinez C. An algorithm to derive a numerical daily dose from unstructured text dosage instructions. Pharmacoepidemiol Drug Saf. 2006 2006 Mar;15(3):161-6.
4. Lau LM, Lam SH. Applying the Desiderata for Controlled Medical Vocabularies to Drug Information Databases. AMIA Annual Symposium Proceedings. 1999:97-101.
5. C. Friedman, G. Hripcsak, W. DuMouchel, S. B.Johnson, and P. D. Clayton. Natural Language Processing in an Operational Clinical Information System. Natural Language Engineering 1 (1):83-108, 1995.