Limited Parsing of Notational Text Visit Notes: Ad-hoc vs. NLP Approaches

Randolph C. Barrows Jr. MD MS, Mihai Busuioc MD, Carol Friedman PhD
Department of Medical Informatics, Columbia University, New York, NY

Abstract

This paper describes the extraction of structured data relevant to glaucoma diagnosis and progression from visit notes typed as "notational text" by ophthalmologists during patient encounters. We compared two text processing systems: a limited pattern matching system called GDP (Glaucoma Dedicated Parser) and MedLEE, a proven natural language processing system which is in routine use encoding findings from chest radiograph and mammogram reports at the NewYork-Presbyterian hospital's Columbia-Presbyterian Center. We also evaluated the use of GDP as a preprocessor program to transform notational text into constructions recognizable by MedLEE. These systems have been evaluated according to their recall and precision in the particular task of processing a corpus of "notational text" documents to extract information related to glaucoma disease.

Introduction

Electronic medical record (EMR) systems approach the problem of data capture from clinicians in one or both of two general ways. The first consists of using domain-specific user interfaces for capturing data in a structured, and potentially code-able, format. The second consists of collecting and storing clinical information in computer systems in text format. "Plain text" capture is often easier and cheaper to implement, but in order to "unlock" the information for sophisticated uses such as computerized decision support and outcomes measurements, it must be structured and coded. Medical language processing (MLP) systems that extract, structure and codify information from textual patient reports have been developed to address this problem [1-9].

A special kind of text in clinical documentation is "notational text", a terse form of written documentation by clinicians that is full of abbreviations and symbols, some of which may be specific to a medical sub-domain, to an institution, or indeed, even a clinician. Statements in notational text are poorly formed according to usual grammatical construction rules. However, domain experts generally have little trouble deciphering notational text due to familiarity and clinical context. For example, the following notational text sample

3/11/98 IPN
SOB & DOE ↓
VSS, AF
CXR ⊕LLL ASD no Δ
WBC 11K
S/B Cx ⊕GPC c/w PC, no GNR
D/C Cef → PCN IV

means: (date of) Intern Progress Note, the patient's shortness of breath and dyspnea on exertion are decreased, the patient's vital signs are stable and the patient is afebrile, a recent new chest xray shows a left lower lobe air space density that is unchanged from the previous radiograph, a recent new white blood cell count is 11,000 cells per cubic milliliter, the patient's sputum and blood cultures are positive for gram positive cocci consistent with pneumococcus, and no gram negative rods have grown, so the plan is to discontinue the cefazolin and then begin penicillin treatment intravenously.

The motivation for our current work came from a need to extract clinical parameters of glaucoma from ophthalmology visit notes written in notational text format, then provide this data as input to an expert system for predicting glaucoma disease progression. Due to a dearth of published experience with notational text processing, we report on the comparative utility of an ad-hoc text processing system for this task, and whether a proven MLP system, effective in understanding transcribed medical text, can be made to perform adequately on notational text.

Background

Glaucoma is a common disease causing irreversible visual field loss. It is strongly associated with high intraocular pressure (IOP), but some persons with "normal" eye pressure become afflicted while others with intraocular hypertension remain spared of visual field loss; thus other factors are at play. One model suggests that certain physiologic parameters determine an individual's intraocular pressure tolerance (IPT) which can mitigate or potentiate the effect of any given IOP in causing glaucoma. One of us (MB) has developed a predictive system based upon this model, and seeks clinical data with which to validate it. Thus we sought clinical parameters related to glaucoma from a large corpus of visit notes created by ophthalmologists. Since these notes were not dictated, but were typed by physicians during routine patient encounters, they were full of abbreviations and symbolic constructions, with a dearth of punctuation. As such, they were more characteristic of a "notational text" than transcribed narrative text. Still, we decided to undertake the challenge of extracting the following clinical parameters from these notes: age, sex, intraocular pressure, cup:disk ratio, visual fields, retinal vascular status, and glaucoma diagnosis. Blood pressure is also an important clinical parameter in this model, but was rarely measured by ophthalmologists, and so was not sought from this corpus of visit notes, but was available from another source.

Since much of human knowledge is recorded in linguistic form, computers that understand natural language (NL) could ease the burden of knowledge acquisition from experts, and automate the input and encoding of information for use in computer systems. Natural language processing (NLP) and understanding depends upon a computational model of human language that uses knowledge about the structure of (the source) language, including what words are (morphological knowledge), how words combine to form sentences (syntactic knowledge), what words mean and how word meanings contribute to sentence meanings (semantic knowledge), and how previous discourse affects current interpretation (discourse knowledge). Also important in natural language is general world knowledge that helps to resolve word and phrase references (pragmatic knowledge), making language so flexible and expressive. Phonetic and phonological knowledge is concerned with how words relate to the sounds by which they are vocally expressed.

NLP systems typically utilize distinct components or processes for each of these different types of language knowledge. Systems often utilize three principal levels of knowledge representation that map from one to another: syntactic processing, concerned with the structural properties of sentences; semantic processing, which computes a logical form that represents a context independent meaning; and contextual processing, which modifies meanings according to their domain of expression. The process that maps a unit of language, such as a sentence, to its syntactic structure and logical form is called the parser. It uses a lexicon, which provides knowledge about words and word meanings, and a grammar, which is a set of rules defining the legal structure of sentences and their parts. Interestingly, for non-medical text documents, the results of a previous evaluation demonstrated that NLP systems utilizing simpler pattern matching algorithms and only limited linguistic knowledge performed well compared to those containing more complex linguistic knowledge [9].

A variety of techniques have been used by MLP systems. Some, such as the LSP [2] system and Ménélas [3], use comprehensive syntactic and semantic knowledge. RECIT [4] uses syntax to recognize the structure of local phrases, and interleaves phrase recognition with semantic knowledge in order to assemble semantically relevant groupings and representations. MedLEE [6,10] relies heavily on general semantic patterns interleaved with some syntax, and also includes knowledge of the structure of the entire sentence. SymText [7] was initially purely semantically driven, and worked by selecting relevant semantic frames associated with semantic information about the words in a sentence and expectations about findings, locations, and conditions. More recent versions integrated syntax into the processing. Other MLP systems use methods that are based on pattern matching and keyword search [1].

MedLEE (**Med**ical **L**anguage **E**xtraction and **E**ncoding) is an MLP system that has been used to extract, structure, and encode clinical information from all chest radiograph and mammogram reports for the past 5 years at the New York-Presbyterian (NYP) hospital's Columbia-Presbyterian Center (formerly Columbia-Presbyterian Medical Center). On average, MedLEE processes about 650 chest radiograph and mammogram reports daily, and the coded data are stored in the hospital's clinical database. Coded data have been used for automated decision-support, as well as formal and ad-hoc analyses [10-11].

MedLEE consists of component software modules, each of which processes the text in some way and generates output used by subsequent components. Each module results in a further "regularization" (formal structuring) of input text without significant loss of information. The first component is the preprocessor, which utilizes tokenization rules to determine word and sentence boundaries, resolve abbreviations, and perform lexical lookup, which finds semantic role-definitions of words and phrases in the sentences. The second component is the parser, which utilizes lexical definitions and grammar rules to determine the structure of a sentence and interpret relationships among sentence elements. A third component is the phrase regularizer, which further structures the target form by formally composing multi-word terms that have been separated in prior component output. Finally, the encoding component maps regularized output to coded vocabulary terms.

MedLEE tries to analyze the structure of an entire sentence using a grammar that consists of patterns of semantic and syntactic categories that are well formed. For example, **finding in bodyloc conj bodyloc** is a well-formed pattern corresponding to sentences such as "pain in arms and legs". If parsing fails, various recovery modes are utilized in order to achieve robustness, each of which is likely to increase sensitivity but at the expense of decreasing specificity and precision. The most specific method is attempted first, and successively less specific methods are used as needed. First, (1) the initial segment is the entire sentence and all words and multi-word phrases must be defined and fitted to a well-formed pattern matching this sentence; failing this, (2) the sentence is segmented at certain types of words or phrases (e.g.: "consistent with") and an attempt is made to recognize each segment independently; failing this, (3) an attempt is made to identify a well-formed pattern for the largest prefix of the segment, which might be successful when the first part of a sentence contains a well-formed pattern but the end does not; failing this, (4) undefined words are skipped and an analysis is attempted starting again with mode (1); and failing this, (5) the first word or phrase in the segment associated with a primary finding (i.e. "infiltrate", "mastectomy", etc.) is identified, after which an attempt is made to recognize the part of the segment starting with the leftmost modifier of the finding and, if analysis fails, recognition is again attempted starting at the next modifier to the right, and continued thus until a successful analysis is obtained. For (5), a modification handles negation, which may have to be distributed over all subsequent segments.

The task of extending a NL processor from its original domain of expertise is generally easier if the new domain is similar (such as another subdomain within Radiology), perhaps requiring only extension of the lexicon. A new domain of application, however, may additionally require new preprocessing rules, grammar patterns and semantic categories [6]. While MedLEE was never designed to process (grammatically malformed) notational texts, we hypothesized that, once abbreviation and symbol recognition was achieved, semantic patterns in notational text phrases would approximate those of well-formed sentences, sans relatively meaningless articles, prepositions, and connectives. Note that a few important exceptions exists, e.g., a ',' often means "and" in the middle of a notational text phrase, and a '/' can mean "and" or "or", etc. Sample content from an ophthalmology visit note follows.

va wc
od 20/ 60 ph 20/50-2
os 20-/100 ph ni stable
p 4-2 reactive ou no rapd
eom full
sle: lla mild blepharitis ou
c.s 1+ papillary rxn ou
k inf spk ou
ac d+q ou
i rr ou no rubeosis
l 2+ ns ou brunescent
ta 14 ou
pp m1/m2.5
c:d 0.3 ou
no bdr
hyperpigmented scar at and inferior to macula os-
unchanged

Methods

A corpus of 12,839 ophthalmology visit notes, described above, was obtained from the NYP clinical data repository. The corpus was obtained in a compacted ASCII format and was expanded so that each encounter note consisted of a sequence of characters (including spaces) separated by a carriage return-line feed sequence.

After ad-hoc analysis of the corpus, a glaucoma-dedicated parser (GDP) was designed and implemented, utilizing pattern matching of words and phrases representative of the clinical parameters sought. The program also makes use of available section identifiers (e.g. "PAST OCULAR HISTORY:", "IMPRESSION:") for context information, and tries to deduce the role of white space as potential phrase and sentence delimiters. After identification, glaucoma-related clinical parameters are extracted together with their values and output to a database in parameter-value tuples, or (optionally) output to a file in short sentence format (e.g. "The IOP in OD is 20.").

One challenge facing our use of MedLEE was its application to a new domain, and another was the lack of punctuation and abundance of (grammatically) poorly formed constructions that characterize notational text. MedLEE's lexicon was extended with definitions of words and phrases representative of the clinical parameters sought. These single and multi-word phrases were semantically categorized (as a **finding**, **location**, **body part**, etc.) and added to the lexicon with a specification of their target form. For example, the construction

wdef(scotomas,cfinding,[scotoma,[quantity,'>1']])

defines "scotomas" as a **finding** with target form "scotoma" and **quantity** greater than one. Common abbreviations, symbolic constructions, and misspellings were similarly defined. MedLEE's performance was tuned by defining new "sections", identified by tags ("Impression:", "sle:", etc.) to provide context sensitivity; by recategorizing the semantic role of some words and phrases; and by preprocessing the input to indicate "sentence" boundaries via a '.' (period). For example, by recategorizing "OD" (ocular dexter = "right eye") and "OS" (ocular sinister = "left eye") from **bodylocation** to **bodymeasurement**, each token was allowed to be associated with a parsed value, such as for visual acuity ("Va"), which the parser assigns automatically. By making "Va" a "section" tag, instances of "OD" and "OS" with associated values were understood as right and left eye visual acuity measurements.

Grammar rules were neither modified nor extended to optimize MedLEE performance for this work.

Results

The corpus of ophthalmology notes was processed by both GDP and MedLEE, each of which generated structured output indicating the clinical parameters found in the visit notes. One hundred (100) notes were randomly selected for manual analysis. One of us (MB) highlighted the desired clinical parameters present in each of these 100 reports, and GDP and MedLEE output was compared. For each clinical parameter sought, an assessment was made whether automatic retrieval was successful (TP) or not (FN), and whether misunderstandings occurred (FP) or not (TN). For each desired clinical parameter, recall (TP/[TP+FN]), precision (TP/[TP+FP]), and accuracy ([TP+TN]/(TP+FP+TN+FN]) were calculated as a measure of system performance. The performance of GDP is summarized as follows:

	Recall	Precision	Accuracy
Age	0.95	0.99	0.94
Sex	0.99	0.97	0.97
IOP	1	0.98	0.99
Cup/disk ratio	1	0.96	0.98
Vascular status	1	0.92	0.97
Visual field	0.96	1	0.99
Diagnosis	0.96	0.89	0.96

MedLEE performance is summarized as follows:

	Recall	Precision	Accuracy
Age	0.88	1	0.89
Sex	1	1	1
IOP	0.85	1	0.89
Cup/disk ratio	0.8	1	0.91
Vascular status	0.8	1	0.94
Visual field	1	1	1
Diagnosis	.94	1	0.99

From the total number of 12839 visit notes, GDP discovered 11,286 values for Age (in 88% of notes), 9468 values for Sex (in 74% of notes), 9979 values for Intraocular Pressure in at least one eye (in 78% of notes), 5338 values for Cup/Disk Ratio in at least one eye (in 42% of notes), 2939 values for Retinal Vascular Status in at least one eye (in 23% of notes), 3175 values for Visual Field in at least one eye (in 25% of notes), and a Glaucoma Diagnosis in 3830 (30%) of the notes. All needed parameters co-occurred in only 267 (2%) of visit notes, but a single patient often has all needed parameters collected over more than one visit, so spanning a set of two or more visit notes, and this latter count was not tallied.

Similar counts for MedLEE's extraction of the clinical parameters were not tallied, but based upon recall measurements they can be estimated as similar, or slightly lower for some of the parameters.

Discussion

For this domain of ophthalmology visit notes, for this text type of "notational text", and for the limited task of extraction of specific clinical parameters (not a general understanding of the text), the pattern-matching and ad-hoc approach of GDP had better recall than the NLP approach, but the NLP approach of MedLEE had better precision. However, the precision and recall of both systems were reasonably acceptable for their intended use.

The fact that MedLEE worked at all for understanding notational text is somewhat remarkable, since it was designed for use on well-formed narrative text. MedLEE's performance in this regard can probably be attributed to its robust recovery mechanisms, discussed above. One observation, not reflected in the above data, was that MedLEE's performance apparently improved significantly when the notational text was preprocessed to indicate probable "sentence" boundaries. The dearth of sentences, as delineated by punctuation, is a typical feature of notational text. Fortunately "phrases", or "segments", with which MedLEE works well, are also characteristic of notational text. Such phrases, however, are usually separated by white space rather than punctuation, so it makes some sense that MedLEE might perform well on notational text once assisted with the segment recognition task.

Another observation was the misinterpretation of some tokens occurring in these ophthalmology visit notes: ones that were homonymous with those used in radiology reports. For instance, MedLEE interpreted the segment "glaucoma suspected started on T1/2 OU" as "**problem:** glaucoma" with "**bodyloc:** first thoracic vertebrae". Here MedLEE recognized "T1/2" as an anatomical reference to thoracic vertebrae, such as discussed in chest xray reports, rather than "**treatment:** Timoptic 0.5% drops in both eyes" as would be correct. This is because MedLEE had previously been trained mostly for use in radiology reports, and our "retraining" for the domain of ophthalmology was incomplete.

As if natural language understanding of narrative text documents by computer systems is not difficult enough, the understanding of notational text documents is perhaps even more difficult due to lack of punctuation and grammar, and frequent use of terse abbreviations and symbols. However, a significant and clinically useful portion of medical documentation exists in notational text format, in progress notes, in sign-out notes, on radiograph jacket covers, and elsewhere. Clinicians who adapt to clinical data entry via computers, particularly text data entry via the keyboard, will take their terse documentation style with them from paper to electronic form, as evidenced here in our corpus of ophthalmology visit notes. Unlocking the information contained in this documentation provides an interesting Informatics challenge for the new millenium.

Conclusion

For our corpus of ophthalmology visit notes, in the "notational text" format typed by physicians, a pattern-matching and ad-hoc approach to structured data extraction had better recall than an NLP approach, but the NLP approach had better precision. However, the precision and recall of both systems were acceptable for their intended use. It is remarkable that a proven NLP system trained for well-formed narrative text had reasonable performance at all. Efforts to unlock the clinical information contained in notational text notes could contribute significantly to the availability of clinical data in computer systems for decision-making, decision support, and process and outcomes analyses.

References

1. Spyns P. Natural language processing in medicine: an overview. Methods of Information in Medicine, Dec 1996; 35(4-5):285-301.

2. Sager N, Lyman M, Buchnall C, Nhan N, Tick L. Natural language processing and the representation of clinical data. JAMIA 1994; 1(2): 142-60.

3. Zweigenbaum P and the Menelas Corsortium. Menlas: an access system for medical records using natural language. Comput Meth Prog Biomed 1994; 45:117-20.

4. Baud RH, Rassinoux AM, Scherrer JR. Natural language processing and semantical representation of medical texts. Methods of Information in Medicine Jun 1992; 31(2): 117-25.

5. Baud RH, Lovis C, Rassinoux AM, Scherrer JR. Morpho-semantic parsing of medical expressions. Proceedings / AMIA Annual Symposium, 1998: 760-4.

6. Friedman C. Towards a comprehensive medical language processing system: methods and issues. Proceedings / AMIA Annual Fall Symposium, 1997: 595-9.

7. Haug P, Koehler S, Lau M, Wang P, Rocha R. Experience with a mixed semantic/syntactic parser. Proceedings / SCAMC 95, 1995: 284-88.

8. Nazarenko A, Zweigenbaum P, Bouaud J, Habert B. Corpus-based identification and refinement of semantic classes. Proceedings / AMIA Annual Fall Symposium: 585-9, 1997.

9. Friedman C, Hripcsak G, Shablinsky I. An evaluation of natural language processing methodologies. Proceedings / AMIA Annual Symposium: 855-9, 1998.

10. Hripcsak G, Friedman C, Alderson P, et al. Unlocking clinical data from narrative reports. Ann Int Med 1995; 122(9): 681-8.

11. Hripcsak G, Knirsch CA, Jain NL, Pablos-Mendez A. Automated tuberculosis detection. JAMIA Sep-Oct 1997; 4(5): 376-81.