A Natural Language Parsing System for Encoding Admitting Diagnoses

Peter J. Haug, Lee Christensen, Mike Gundersen, Brenda Clemons, Spence Koehler, Kay Bauer

LDS Hospital and the University of Utah

ABSTRACT

Free-text or natural language documents make up an increasing part of the computerized medical record. While they do provide accessible clinical information to health care personnel, they fail to support processes that require clinical data coded according to a shared lexicon and data structure. We have developed a natural language parser that converts free-text admitting diagnoses into a coded form. This application has proven acceptably accurate in the experimental laboratory to warrant a test in the target clinical environment. Here we describe an approach to moving this research application into a production environment where it can contribute to the efforts of the Health Information Services Department. This transition is essential if the products of natural language understanding research are to contribute to patient care in a routine and sustainable way.

INTRODUCTION

Medical information systems are designed to capture and manipulate large amounts of clinical data. These data can take a number of forms. The largest portion in most modern systems takes one of two forms, either free-text or coded data. Free-text is typically the information that is dictated by a care giver and typed into a computer by a transcriptionist. It is frequently referred to as natural language data. Coded data is information that is typically entered in a structured way and stored according to a data dictionary and a pre-defined storage structure.

Natural language documents can be shown on a computer screen or printed and are typically understood easily by the care givers who read them. However, the data is largely inaccessible to computer programs that manipulate medical information symbolically for research, medical decision making, quality assurance initiatives, and the management of the medical enterprises.

In contrast, data in a coded form can be used in research, decision support, quality assurance, analyses done for management purposes and in a variety of focused reports that combine information from multiple sources to flexibly address a particular information need.

Natural language is common in medical systems and is becoming more common. Not only is dictation and transcription widespread, but new technologies will soon make free-text documents even more easy and inexpensive to produce. Before the turn of the century, computer systems that convert speech to text will supplant a significant number of medical transcriptionists.

In order to make coded data available in a setting where a large subset of the information will reside in natural language documents, a technology called natural language understanding (NLU) is required. This technology allows a computer system to "read" free-text documents, to convert the language in these documents to predefined concepts, and to capture these concepts in a coded form in a medical data base.

Several groups have evaluated techniques for NLU. Well known among these groups is the Linguistic String Project which has developed a series of tools for analyzing medical text [[1],[2]]. Gabrielli has described a system for encoding discharge summaries for quality assurance [[3]].

Chest x-ray reports appear to have a special appeal. Zingmond has applied a semantic encoding tool to these reports to recognize abnormalities that should receive follow-up [[4]] and Friedman has studied techniques for encoding interpretations found in these reports [[5]].

The techniques described by Friedman were the basis for efforts by Hripsak et. al. to study the potential for NLU systems to abstract findings from chest x-ray reports that are relevant to automated medical decision support[[6]]. This group demonstrated that, for encoding a set of six clinical conditions often described on these x-rays, the parsing system described by Friedman was indistinguishable from a group of 12 physicians and was superior to six lay observers and 3 other computer methods.

In recent years, a research effort at the LDS Hospital in Salt Lake City has focused on developing NLU technology for the x-ray department. This effort has resulted in a NLU tool set called SymText[7,8] This tool set is capable of processing natural language documents collected in the HELP hospital information system[9] used in this facility.

In an effort to determine how general this semantically-based technology was, we determined to test the basic SymText tool set in the Health Information Services Department. Here the challenge is the encoding of diagnostic information from the free-text, admitting diagnoses input by registration personnel. These reasons for admission are entered at the time of patient registration and are typically short sentences or simple phrases.

The incentive to attack this problem is the need to have a coded summary of the reason for admission in the HELP data base early in a patient's stay. The natural language reasons input by the registration clerks have been routinely manually encoded by ICD coders from the Health Information Services Department (HIS). Their goal is to encode this information within 24-48 hours of registration. Because of the effort involved in the expeditious capture of this coded information and the time delay involved in waiting for a human coder, we have developed automated approaches for encoding the admit diagnosis. The application produced is called the Automatic Admit Diagnosis Encoding System (AADES)[10].

In a study of this application, the computerized NLU system proved accurate enough to lead us to design the group of processes necessary to provide a production system for early encoding of the admitting diagnosis. Below we describe the components of this system. Its goals are the timely and accurate encoding of each patient's admitting diagnosis in an active hospital information system.

METHODS

The goal of the system described here is to convert a free-text statement of the reason for admitting a patient into one of more than 450 different ICD diagnoses. The AADES system is not currently capable of complete accuracy in this endeavor. We have therefore chosen to provide a set of tools that will give as high an immediate accuracy as possible, and will allow easy correction of coding errors by expert coders in HIS. The goal is a system that can recognize when it is likely to fail and, in these cases, not store results. It must assist the coders in correcting coding failures whenever possible. And it must facilitate improvement of its own accuracy by providing an analysis of its errors including a complete assessment of its failures, as well as a streamlined tool for retraining the system using the results of a review by the HIS coders.

Intelligent Failures

The admitting diagnoses provided by this system are designed to be available within minutes of the input of a free-text "reason for admission". The system cannot be 100% accurate for a variety of reasons. The principle reasons are 1) diagnoses which it has never seen may appear, and 2) misspelling errors are made commonly in the entry of this data. Currently, the system has no facility for unattended correction of spelling errors.

For the potential erroneous ICD-9 codes that the system may propose, a simple strategy significantly reduces the number that are entered into the HELP data base. The NLU model which we have chosen to implement is a probabilistic one. This allows us to set a threshold for accepting the encoding of a admitting diagnosis. In our original testing, 16% of the coded results produced fell below this level. These ICD codes are not stored in the information system.

The codes that are stored in the system are divided into two subgroups at the time of storage. These are: 1) Those codes that are stored provisionally. These codes are rapidly accessible to applications that need to know the admitting diagnosis to participate in the decision support process. However, they are marked as provisional and can be reviewed early by HIS personnel. 2) Those codes that can be accepted automatically and stored without subsequent review.

This second set of codes are chosen based on two criteria. First, their probability must exceed the threshold for provisional acceptance and second the ICD code involved must have a record of at least 10 consecutive successful encodings by the NLU system. An automatically encoded admitting diagnosis that meets these criteria is included in a list of codes that receive limited review. One in ten of these will be submitted to HIS coders for critical review. The remainder will be stored without further HIS involvement.

Tools for Review

The Figure 1 is a diagram of the flow of information through the production system for managing admitting diagnoses. The process begins with the diversion of the free-text admitting reason to the AADES system. Here the NLU process determines an admit diagnosis and its associated probability. Based on this probability, the parsing result either succeeds or fails. The free-text associated with failed parses is diverted to a queue from which it can be manually encoded. The results of the manual process are stored in the HELP system and added to the training data set used to update the capabilities of the parser on a routine basis.

When the results of the automated coding process exceed the system's provisional acceptance threshold they are divided into two subsets. One of these subsets contains records that are stored provisionally in the HELP system. They are reviewed at a latter date by HIS coders and the results of this review are used to update the information in the HELP system. If the original result was correct, the label for that data element is altered from provisional to final. Otherwise, the corrected final result is added to the patient record.

The second set of admit diagnosis codes have passed both the provisional acceptance threshold and are among the codes with history of accuracy as described above. These ICD codes can be stored without subsequent review. However, one in ten of the automatic acceptance records for each ICD code are routed to the provisional group to allow a continuing monitor of the accuracy of the NLU system in this subset of records.

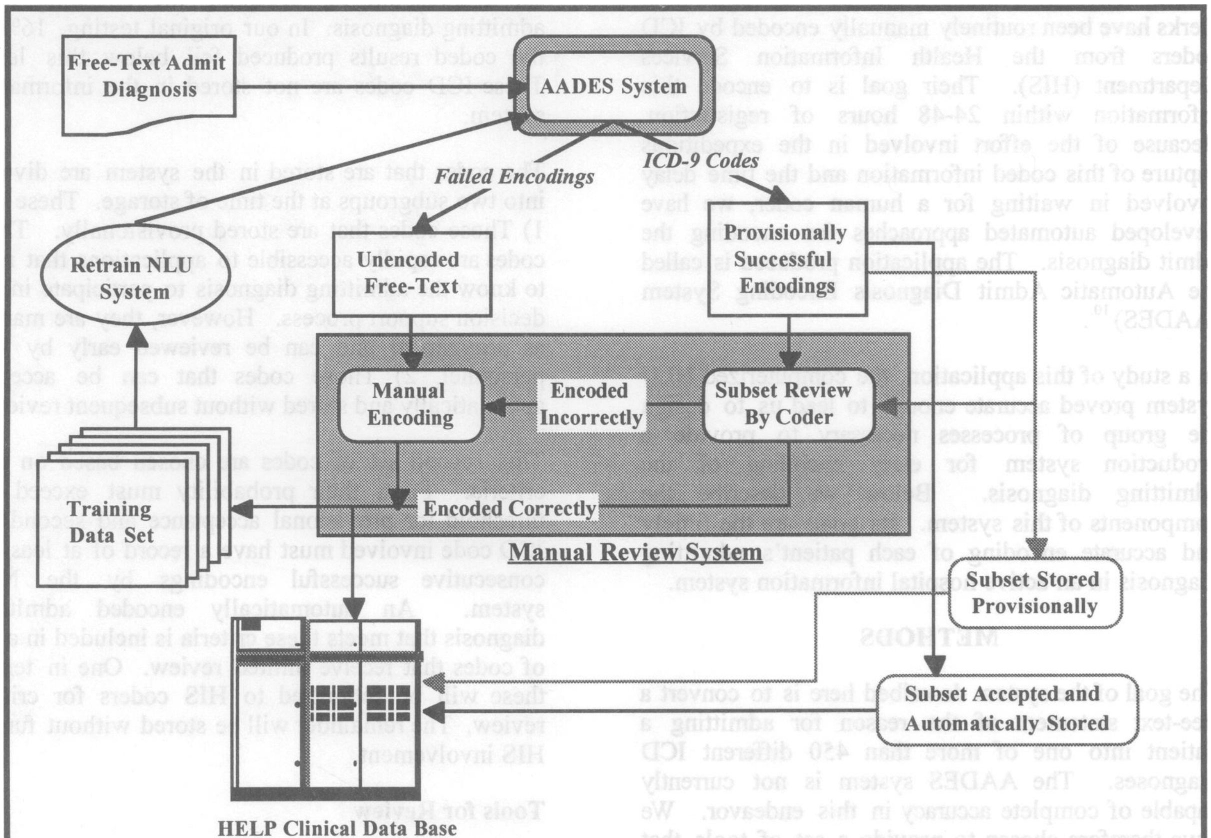

Figure 1: Review and editing model for the automated encoding system. Initially all of the computer coded admit diagnoses will be placed in the "store provisionally" group and reviewed by HIS personnel. As subsets with high coding accuracy are identified they can be accepted and stored without routine review.

The review application has two features that help to streamline the process of editing the results of the NLU application. The first is based on the probabilistic character of the parser. When a HIS coder reviews either the coding failures or the provisionally coded results, the review application shows the top three most likely encodings for each free-text admitting reason. Often, correcting an erroneous parse involves simply choosing the proper alternative from this list. In addition, the review application contains a feature that allows the planned alteration of a group of ICD results. This feature is particularly useful in these cases where HCFA mandates a global change in the way certain conditions are coded.

This feature simply allows the reviewer to identify codes that are not used anymore and instruct the system to change all future instances of these codes before presenting them for review. The results are propagated throughout the remainder of the results being reviewed and are added to the training set to be incorporated into the knowledge base of the NLU system during the next training session.

Metrics

The results of this process are summarized in a set of statistics designed to allow regular tracking of the systems accuracy. Figure 2 displays examples of the output of this accuracy monitoring system.

The first example in Figure 2 is an error report designed to indicate the overall accuracy of the AADES System. The second example is a breakdown of the errors produced by the system. This report is used to plan improvements to the underlying algorithms, to determine the adequacy of the training set, and to monitor the overall effectiveness of this system in meeting expectations of increased accuracy in the coded admitting diagnoses and efficiency in the work of HIS personnel.

Error Report	
# Sentences Parsed:	**469**
Percent Correct:	64.6
Percent Incorrectly Parsed:	19.6
Percent Unparsed:	11.9
Percent Unparsable	3.8
Patients Auto-encoded:	75
Codes in Auto-encoding List:	2

Error Analysis	
Number of Parses Corrected:	**85**
Error Types	
New Diagnoses (%):	26.8
Misspellings (%):	18.5
Incorrect Syntactic Processing (%):	15.6
Incorrect Semantic Processing (%):	13.9
Missing Synonym (%):	8.4
Error in Prior Training Set (%):	3.7
Other Errors (%):	13.7

Figure 2: Examples of two reports describing the behavior of the AADES System. Detailed subsets from these reports have been collapsed into more informative supersets.

System Retraining

A side-effect of the review and editing done by the HIS coders is a collection of corrected ICD-9 codes each associated with the phrase or sentence from which it was derived. This is the beginning of a training set from which the accuracy of the system can be improved.

We have developed a set of tools that allow these error records to be used as training cases to enhance the accuracy of AADES. In order to create new training cases from the error set, personnel familiar with the requirements of AADES review the error records and determine the best way to present them to the training application. After this preprocessing, the new training cases are added to the old training set and a new version of AADES's knowledge base is created. In the next set of error reports, we look for improvements in the behaviors that previously resulted in errors.

The process of turning the system errors into new training cases is the most labor intensive part of this process. If the cycle of weekly retraining we have instituted is effective, we expect the cost of this training process to drop steadily as the accuracy of the system rises.

DISCUSSION

Computer-based natural language understanding is a challenging area. The flexibility of human understanding appears impossible to emulate completely. Whenever a new registration clerk begins to enter his/her version of the admitting reason into the computer, a new set of linguistic nuances and a new set of misspellings and abbreviations appear in the free-text that this system interprets. Usually these variations are well within the ability of a medically knowledgeable human to disambiguate, but they frequently prove uninterpretable for AADES.

One of the long-term goals of the system described here is to provide a test bed to determine to what degree the continuous training of a NLU system can compensate for this natural language variability. In addition, we hope to be able to ask which alterations to the underlying models might also extend to the accuracy of this system. Already it is clear that a tool for the unattended correction of spelling errors will be essential to the long-term success of NLU in this environment.

A solution to the general problem of capturing coded data is often mentioned in this context. Natural language could be largely eliminated from the medical record. Good applications for structured data collection could supplant many of the programs that now allow the storage of free-text. Proper design and dissemination of these processes could make natural language a rarity in the medical record and could remove the question of mechanisms to interpret it from consideration.

However, natural language, both dictated and written are unlikely to disappear easily from our medical computing systems. These forms of communication are too natively human to be supplanted by clever computer interfaces. Ultimately, we will be required to make computers more human-like rather than humans more computer-like.

The processes described in this report are focused in the area of diagnostic encoding. The emphasis is on the admitting process. However, the challenges are similar in other realms where the value of free-text information would be magnified by converting it to a coded form. We are presently exploring similar processes for encoding information embedded in radiology reports. Here the value of coded information for decision support, quality assurance, and enterprise management are also clear. We believe that applications for encoding natural language will be a standard part of future medical information management and that tools for the routine maintenance of NLU applications will be standard features in these systems.

Supported in part by Grant #LM05323 from the National Institute of Medicine and Grant #HL53427 from the National Heart, Lung & Blood Institute

References:

1. Sager N, Friedman C, Lyman MS. *Medical Language Processing: Computer Management of Narrative Data.* Addison-Wesley, Menlo Park, CA (1987).

2. Sager N, Margaret L, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. JAMIA (1994); 1:142-160.

3. Gabrielli ER. Computer assisted assessment of patient care in the hospital. J Med Syst(1988); 12:135.

4. Zingmond D, Lenert LA. Monitoring free-text data using medical language processing. Comp Biomed Res (1993); 26:467-481.

5. Friedman C, Alderson PO, Austin JHM, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. JAMIA (1994); 1:161-174.

6. Hripsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking Clinical Data from Narrative Reports: A Study of Natural Language Processing. Ann Intern Med (1995); 122:681-688.

7. Haug PJ, Koehler S, Lau LM, Wang P, Rocha R, Huff S. A Natural Language Understanding System Combining Syntactic and Semantic Techniques. Eighteenth Annual Symposium on Computer Applications in Medical Care (1994), pp 247-251.

8. Haug PJ, Koehler S, Lau LM, Wang P, Rocha R and Huff SM. Experience with a Mixed Semantic/Syntactic Parser. Nineteenth Annual Symposium on Computer Applications in Medical Care (1995), pp 284-288.

9. Kuperman GJ, Gardner RM, Pryor TA. *HELP: A Dynamic Hospital Information System.* (1991) Springer-Verlag, New York.

10. Gundersen ML, Haug PJ, Pryor TA, Van Bree R, Koehler S, Bauer K, Clemons B. Development and Evaluation of a Computerized Admission Diagnoses Encoding System. Computers and Biomedical Research (1996); 29, 351-372.