Automatic Identification of Pneumonia Related Concepts on Chest x-Ray Reports

Marcelo Fiszman MD, Wendy W. Chapman, Scott R. Evans Ph.D., Peter J. Haug MD

Department of Medical Informatics, LDS Hospital, The University of Utah

ABSTRACT

A medical language processing system called SymText, two other automated methods, and a lay person were compared against an internal medicine resident for their ability to identify pneumonia related concepts on chest x-ray reports. Sensitivity (recall), specificity, and positive predictive value (precision) are reported with respect to an independent panel of physicians. Overall the performance of SymText was similar to the physician and superior to the other methods. The automatic encoding of pneumonia concepts will support clinical research, decision making, computerized clinical protocols, and quality assurance in a radiology department.

INTRODUCTION

Hospital information systems and clinical data repositories have been used to improve the quality and reduce costs associated with health care. One example of such system, the Antibiotic Assistant, has been in production on the HELP system[1] at LDS Hospital for 10 years[2]. A recent evaluation showed that this system is able to reduce the length of patient stay and also reduce costs associated with antibiotic prescriptions[3]. To function properly, these systems depend on coded data present in the electronic medical record of the patient. Most of the laboratory data is encoded with enough granularity to be adequately utilized. However, much of the clinical information of the patient - history, physical, and procedural reports - are usually stored as free-text data and are completely inaccessible to other applications.

In order to make appropriate antibiotic recommendations, one module of the Antibiotic Assistant at LDS hospital checks for the presence of concepts related to acute bacterial pneumonia (i.e., infiltrates, aspiration, pneumonia, etc.) on chest x-rays reports. Because those reports are stored in the HELP system as free-text reports this information is extracted with the help of a key word search program (AACKS - Antibiotic Assistant Complex Keyword Search). The extraction of this type of information from chest x-ray reports is not trivial. The problems come from the different ways used to express the same concepts (synonymy), grammatical ambiguities, and negation distribution (like in the sentence "There is no evidence of atelectasis, pulmonary contusion, and consolidation". In this case is important to recognize the absence of all three concepts, not only the first). Therefore, we decided to compare the performance of AACKS with a medical language processing system that we are developing at LDS hospital.

Medical language processing (MLP) techniques have been used to analyze free text reports in order to extract relevant clinical information to be stored in a central repository[4, 5, 6]. A research effort at LDS Hospital in Salt Lake City has focused on developing an MLP system for chest x-ray reports. This effort has resulted in a tool called SymText[7, 8]. In an attempt to determine how general SymText is, it was tested for its ability to encode free-text admitting diagnoses to produce ICD-9 codes[9]. The system proved to be accurate enough be put into production in the Health Information Service Department[9].

The objective of this study is to test whether SymText is able to identify acute bacterial pneumonia related concepts in chest x-ray reports. SymText's performance is compared against AACKS, a lay person, and a simple key word search (SKWS). Our hypothesis is that SymText can extract concepts related to pneumonia from chest x-ray reports as well as a physician and better than the other methods. For this reason, SymText and the other methods are compared to an internal medicine resident.

The extraction and coding of this type of information is important not only for the antibiotic assistant program, but also for medical decision support, outcome studies, medical, epidemiology, etc.

1091—8280/99/$5.00 © 1999 AMIA, Inc.

METHODS

LDS Hospital is a 500-bed, private, tertiary care facility located in Salt Lake City, Utah. Approximately, 40,000 chest x-ray reports are performed in one year at this hospital. The results are stored as free text reports in the HELP system database. For this study, we randomly selected 150 reports from patients that had a primary discharge diagnosis of pneumonia and 148 reports from patients with primary discharge diagnosis different from pneumonia (n = 298 reports). We adopted this procedure to increase the prevalence of pneumonia in our sample.

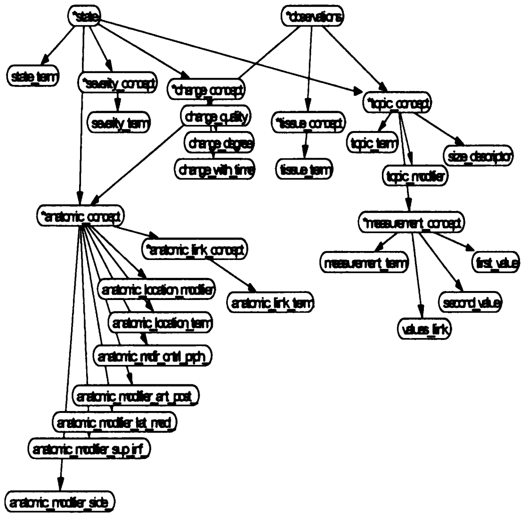

Figure 1- Bayesian Network for radiographic findings

SymText has a syntactic and a semantic component. Syntactic information is implemented as a set of augmented transition network grammars[10](ATN) followed by the application of a transformational grammar. The semantic knowledge of the system is stored in three bayesian networks[11]. The ATN and the transformation grammar have access to the semantic information as they operate. The first bayesian network models the appliances (devices) that are frequently described in the report (i.e., swanz-ganz, lines, nasogastric tubes, etc). The second bayesian network represents radiographic findings (infiltrates, pleural effusions, mediastinal widening, etc). The third network models the diseases that are described in the reports (i.e., pneumonia, congestive heart failure, atelectasis, etc). Currently, the last two networks represent 76 findings and 89 diseases, respectively. However for this study, driven by the kind of data used by the Antibiotic Assistant, we are only interested in pneumonia related concepts.

Instantiated Event:
1001 *observations : ***localized upper lobe infiltrate*** (0.888649)
1002 *state : ***present*** (0.989832)
1003 state term : null (0.966054)
1004 *topic concept : *poorly-marginated opacity (infiltrate) (0.877889)
1005 topic term : opacity~n (1.0)
1006 topic modifier : infiltrative~adj (1.0)
1007 *measurement concept : *null (0.999086)
1008 measurement term : null (0.990915)
1009 first value : null (0.998183)
1010 second value : null (0.999999)
1011 values link : null (0.999999)
1012 size descriptor : null (0.999507)
1013 *tissue concept : *lung parenchyma (0.906629)
1014 tissue term : alveolar~adj (1.0)
1015 *severity concept : *high severity (0.893566)
1016 severity term : dense (1.0)
1017 *anatomic concept : *right upper lobe (0.999994)
1018 *anatomic link concept : *involving (1.0)
1019 anatomic link term : in (1.0)
1020 anatomic location term : lobe~n (1.0)
1021 anatomic location modifier : null (0.999864)
1022 anatomic modifier side : right (1.0)
1023 anatomic modifier superior/inferior : upper (1.0)
1024 anatomic modifier lateral/medial : null (0.999993)
1025 anatomic modifier anterior/posterior : null (0.999989)
1026 anatomic modifier central/peripheral : null (0.955543)
1027 *change concept : *null (0.569735)
1028 change with time : null (0.567828)
1029 change degree : null (0.904206)
1030 change quality : null (0.92862)

Figure 2 - Complete instantiation for the sentence "dense infiltrative opacity in the right upper lobe"

SymText tries to make an interpretation for every sentence in the report using the hierarchical model implied in the Bayesian networks as template for the words and concepts. For instance, Figure 1 represents the network that models the radiographic findings and Figure 2 the complete instantiation for the utterance "dense infiltrative opacity in the right upper lobe" structured for storage in a database. Note that nodes represented with an asterisk (*) represent higher level concepts resulting from categorization and abstraction (through probability propagation) from terms represented in lower level nodes

SymText, AACKS, SKWS, the internal medicine resident, and the lay person were tested for their ability to identify the following concepts from the chest x-ray reports: alveolar infiltrate compatible with pneumonia, aspiration, and pneumonia. The presence was marked with 1 and the absence with 0. The subjects were also required to make a decision whether the whole

	Physician											
	Pneumonia			Aspiration			Infiltrate			Support Pneumonia		
	P	R	S	P	R	S	P	R	S	P	R	S
SymText	0.277	0.143	0.433	1.000	0.294	1.000	0.662	**<0.001**	0.883	0.518	0.037	0.452
AACKS	**<0.001**	0.966	**<0.001**	1.000	**<0.01**	1.000	0.079	**<0.001**	0.033	0.820	**<0.001**	0.863
SKWS	**<0.005**	0.306	**<0.005**	1.000	0.118	1.000	**<0.001**	0.702	**<0.001**	**<0.001**	0.703	**<0.001**
Lay	**<0.01**	**<0.001**	0.018	0.094	**<0.001**	0.156	0.991	**<0.001**	**<0.005**	**<0.001**	**<0.001**	**<0.001**

Table 2 - Matrix of p values from multiple comparisons between the physician and the other subjects.
Values in bold represent statistical significance after Bonferroni correction.
P = Precision (Positive Predictive Value); R = Recall (Sensitivity); S = Specificity
AACKS = Antibiotic Assistant Complex Keyword Search
SKWS = Simple Key Word Search

report supports (1) or does not support (0) acute bacterial pneumonia as a

disease in the patient. A different paper[12] studies several different methodologies that allow automatic methods like SymText to make this type of inference. In this study, a simple ruled based system that was empirically developed (did not require training) was applied to the output of all the automatic methods (SymText, AACKS, SKWS).

The gold standard was established according to the following procedure: an independent panel of two internal medicine physicians read the 298 reports and decided on the presence or absence of the concepts and inference described above (ie infiltrate, aspiration, pneumonia, and report supports pneumonia). Disagreements between the two physicians were resolved by an arbitrator (a radiologist with more than 10 years of practice).

We calculated sensitivity (recall), positive predictive value (precision), and specificity on identifying the pneumonia related concepts for each subject in the study. The differences were tested using a Z proportion test with Bonferroni correction for multiple comparisons[13].

RESULTS

		Physician (95% CI)	N	SymText (95% CI)	N	AACKS (95% CI)	N	SKWS (95% CI)	N	Lay (95% CI)	N
Pneumonia	Precision	**0.98** (0.94-1.00)	53	**0.94** (0.88-1.00)	52	**0.79** (0.69-0.89)	66	**0.83** (0.74-0.93)	60	**0.82** (0.71-0.93)	45
	Recall	**0.98** (0.94-1.00)	53	**0.92** (0.85-0.99)	53	**0.98** (0.94-1.00)	53	**0.94** (0.88-1.00)	53	**0.69** (0.57-0.82)	53
	Specificity	**0.99** (0.99-1.00)	245	**0.99** (0.96-1.00)	245	**0.94** (0.91-0.97)	245	**0.95** (0.93-0.98)	245	**0.97** (0.95-0.99)	245
Aspiration	Precision	**1.00**	11	**1.00**	10	**1.00**	7	**1.00**	9	**0.71** (0.38-1.00)	7
	Recall	**1.00**	11	**0.91** (0.74-1.00)	11	**0.63** (0.35-0.92)	11	**0.81** (0.59-1.00)	11	**0.45** (0.16-0.75)	11
	Specificity	**1.00**	287	**1.00**	287	**1.00**	287	**1.00**	287	**0.99** (0.98-1.00)	287
Infiltrate	Precision	**0.94** (0.91-0.98)	172	**0.93** (0.89-0.97)	156	**0.98** (0.96-1.00)	127	**0.81** (0.76-0.87)	198	**0.94** (0.83-1.00)	17
	Recall	**0.98** (0.96-1.00)	165	**0.88** (0.83-0.93)	165	**0.75** (0.68-0.82)	165	**0.97** (0.95-1.00)	165	**0.09** (0.05-0.14)	165
	Specificity	**0.92** (0.88-0.97)	133	**0.92** (0.87-0.97)	133	**0.98** (0.96-1.00)	133	**0.72** (0.65-0.80)	133	**0.99** (0.98-1.00)	133
Support	Precision	**0.95** (0.93-0.99)	179	**0.94** (0.90-0.98)	175	**0.95** (0.92-0.98)	157	**0.82** (0.78-0.88)	205	**0.77** (0.70-0.85)	112
Pneumonia	Recall	**0.98** (0.96-1.00)	174	**0.94** (0.90-0.98)	174	**0.85** (0.80-0.90)	174	**0.97** (0.95-1.00)	174	**0.50** (0.43-0.57)	174
	Specificity	**0.93** (0.89-0.98)	124	**0.91** (0.86-0.96)	124	**0.93** (0.89-0.97)	124	**0.71** (0.64-0.80)	124	**0.79** (0.73-0.87)	124

Table 1 - performance measures for the physician, SymText, the automatic methods and the lay person.
N = Number of reports where the performance measure was calculated.

Table 1 shows the performance measures for all the subjects in the study. Table 2 is matrix of multiple comparisons between the physician and the four methods. In Figure 3, specificities and sensitivities for all subjects are plotted on the inference support pneumonia.

The results demonstrate that the only difference between the physician and SymText was in recall for the concept infiltrate (physician = 98.2% and SymText = 88%). On the other hand, The other methods tend to differ from the physician in several performance measures (see table 2). In recall for the concept infiltrate, SymText was worse than the physician, but still was statistically better than AACKS (88% versus 75% with $p < 0.005$).

DISCUSSION

The performance measures indicate that SymText is closer to the physician than all the other methods in this study. Overall, there was a significant difference between the physician and all the other methods except SymText, as expressed by the comparative matrix of table 2.

We can abstract from figure 3 that, in general, AACKS was more specific than sensitive and SKWS had the opposite tendency. But none of them demonstrated the balance between specificity and sensitivity exhibited by SymText in this study. The lay person was clearly the worst method of all.

Our study design is very similar to a previous evaluation of a natural language processor published by Hripcsak et al[14]. However, in that study six clinical conditions were tested after the output of the parser was interpreted by medical decision logic module. In this study, we measure the performance of SymText on three actual concepts from the reports and the resulting inference from those concepts on acute bacterial pneumonia. Given the different objectives, it is hard to compare the performance measures of both of these studies.

Our study has a couple of limitations. We compared the performance of all the methods with only one physician, when it is known that physicians tend to disagree with each other when making interpretations about patient data.[15] The other limitation is related to the sampling methodology for selecting the reports. The sample used in this study is not a random selection of reports from the general population of patients. We tried to increase the prevalence of pneumonia in our sample by selecting some reports from patients with known discharge diagnosis of pneumonia. Therefore, we may not be able to generalize our results to the real population. However, this is a pilot study, and we are planning to address these problems in a more complete evaluation of the system's performance.

Several applications could benefit from having pneumonia related concepts coded in a database. The first and obvious one is the Antibiotic Assistant program. SymText is more sensitive than AACKS without significant loss of specificity. Therefore, substituting SymText for AACKS would probably improve the overall antibiotic recommendations, although this affirmation has to be tested in the whole context of the Antibiotic Assistant program. Another possible application is to help identifying pneumonia cases for computerized clinical protocols. A decision support application for pneumonia could also benefit from coded information coming from the chest x-ray report. Finally, a quality assurance study on diagnostic interpretations of radiologist could compare pneumonia interpretations with outcome data from other sources of the hospital information system.

In conclusion, we have studied the ability of a medical language processing to identify pneumonia related concepts in chest x-ray reports. The performance of the systems was similar to the physician and superior to the other methods. This automatic encoding supports clinical research, decision making, computerized clinical protocols, and quality assurance.

Acknowledgments

We would like to thank Alexandra Pagel Eidelwein, Scott Evans, Lee Christiansen, Dominik Aronsky, and Robin Webber for their collaboration in this project.

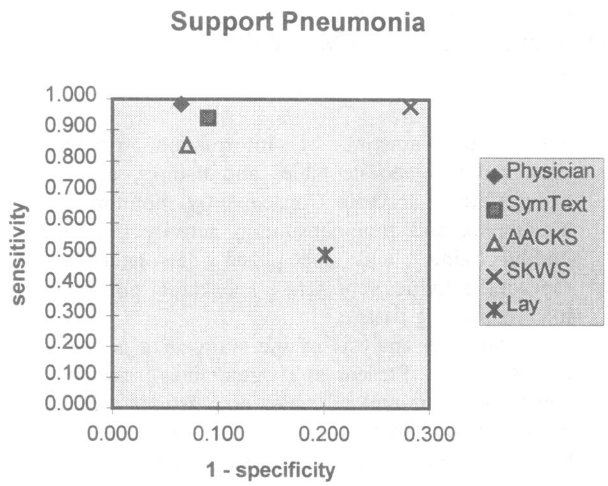

Figure 3 - 1-specificity and sensitivity plotted on ROC axes.

Supported in part by Grant #LM06539 from the National Library of Medicine and Grant #HL53427 from the National Heart, Lung & Blood Institute.

The first author is supported by a scholarship from Research Foundation of the State of Sao Paulo (FAPESP - #96/3922-9).

References

1. Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP system. J Med Syst. 1983, 7:87-102.
2. Pestonik SL, Evans RS, Burke JP, Gardner RM, Classen DC. Therapeutic antibiotic monitoring: surveillance using a computerized expert system. Am J Med. 1990, 88:43-48.
3. Evans RS, et al. A Computer-Assisted management program for antibiotics and other antiinfective agents. NEJM. 1998, 338(4):232-238.
4. Sager N, Lyman M, Bucknall C, et al. Natural language processing and representation of clinical data. JAMIA. 1994, 1:142-160.
5. Friedman C, Cimino JJ, Johnson SB. A schema for representing medical language applied to clinical radiology. JAMIA. 1994, 1:233-248.
6. Zigmond D, Lenert LA. Monitoring free-text data using medical language processing. Comp Biomed Res. 1993. 26:467-481.
7. Haug PJ, Koehler S, Rocha R, et al. A natural language understanding system combining syntactic and semantic techniques. 18th Annual Symposium on Computer Applications in Medical Care. 1994, 247-251.
8. Haug PJ, Koehler S, Rocha R, et al. Experience with a mixed Semantic/Syntactic parser. 19th Annual Symposium on Computer Applications in Medical Care. 1995, 284-288.
9. Gundersen ML, Haug PJ, Pryor A, et al. Development and evaluation of a computerized admission diagnosis and encoded system. Comp Biom Res. 1996, 29:351-372.
10. Allen J. Augmented transition networks in: Natural Language Processing. 2nd edition Benjamin/Cummings Publishing Company, 1994.
11. Pearl J. Probabilistic rasoning in intelligent systems: networks of plausible inference. Morgan Kufmann, 1988.
12. Chapman WW, Haug PJ. Identifying chest x-ray reports that support pneumonia using machine learning and natural language processing. Submitted to AMIA 99.proceedings.
13. Zar JH. Biostatistical Analysis. Prentice Hall Inc, 1974.
14. Hrippcsak G, Friedman C, Alderson P, et al. Unlocking clinical data from narrative reports: A study of natural language processing. Ann Intern Med. 1995, 122:681-688.
15. Yerushalmy J. The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. Rad Clin North Amer. 1969, 7(3):381-392