MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

Identifying Respiratory Findings in Emergency Department Reports for Biosurveillance using MetaMap

Wendy W Chapman[a], Marcelo Fiszman[b], John N Dowling[a], Brian E Chapman[c], Thomas C Rindflesch[b]

[a] *RODS Laboratory, Center for Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA*
[b] *National Library of Medicine, Bethesda, MD, USA*
[c] *Department of Radiology, University of Pittsburgh, Pittsburgh, PA, USA*

Abstract

Clinical conditions described in patients' dictated reports are necessary for automated detection of patients with respiratory illnesses such as inhalational anthrax and pneumonia. We applied MetaMap to emergency department reports to extract a set of 71 clinical conditions relevant to detection of a lower respiratory outbreak. We indexed UMLS terms in emergency department reports with MetaMap, filtered the indexed output with a specialized lexicon of UMLS terms for the domain, and mapped the clinical conditions of interest to concepts in the lexicon. We compared MetaMap's ability to accurately identify the conditions against a physician's manual annotations and evaluated incorrectly indexed features to determine what additional processing is necessary.

MetaMap identified the clinical conditions with a recall of 0.72 and a precision of 0.56. Necessary processing beyond MetaMap's indexing includes finding validation, temporal discrimination, anatomic location discrimination, finding-disease discrimination, and contextual inference. Successful identification of clinical conditions in an emergency department report with MetaMap requires processing techniques specific to the clinical question of interest.

Keywords:

Natural Language Processing, Information Extraction, Biosurveillance, Disease Outbreaks

Introduction

The recent Severe Acute Respiratory Syndrome (SARS) outbreak [1, 2] highlights the need for detailed patient-specific data for biosurveillance. Case definitions of SARS and other potentially infectious respiratory diseases include symptoms and findings such as cough, fever, and air space consolidation that can generally only be found in medical patient records stored in free-text format.

We describe our experience applying an indexing application developed to index biomedical text to UMLS concepts to the task of automatically identifying cardiopulmonary clinical conditions from Emergency Department (ED) reports. We report the performance of the indexing application and describe additional procedures needed for accurate detection of findings from clinical reports.

We implemented an indexing application called MetaMap [3] that was created at the National Library of Medicine and is available for public use (http://skr.nlm.nih.gov). MetaMap performs a shallow parse on a sentence, identifying simple noun, verb, and prepositional phrases. The phrases are normalized for inflectional and derivational variation and are mapped to concepts in the UMLS Metathesaurus [4, 5] . For example, from the noun phrase "severe chest pain" MetaMap generates the UMLS concepts "severe (C0205082)" and "chest pain (C0008031)".

Materials and Methods

Our goal was to use MetaMap to automatically identify from ED reports any of 71 clinical conditions potentially informative for determination of acute lower respiratory syndrome (respiratory features). Applying MetaMap to the task comprised three procedures, shown in Figure 1. First, MetaMap indexed UMLS concepts in ED reports. Second, the indexed UMLS concepts were filtered through a domain lexicon manually compiled from a subset of the Metathesaurus. Third, indexed UMLS concepts were mapped to relevant respiratory features. As shown in Figure 2, the output of the indexing system is an annotated report that identifies individual instances of respiratory features described as occurring at or around the time of the patient's visit to the ED.

We knew at the onset of this project that MetaMap, which has often been applied to indexing UMLS concepts from the biomedical literature, would not be sufficient in and of itself for generating the target output shown in Figure 2. Our aims for this study were twofold: (1) perform an initial evaluation of MetaMap's ability to index the relevant conditions and (2) learn what types of additional procedures are necessary for accurate identification of the respiratory features from ED reports.

Respiratory Features We employed an iterative process involving several physicians to generate a list of clinical conditions potentially helpful in detecting a lower respiratory syndrome. The final list of respiratory features contains 71 clinical conditions, including risk factors for respiratory illness (e.g., HIV/AIDS, pneumonia history), conditions that may indicate or occur with a respiratory illness (e.g., cough, shortness of breath, wheezing, pneumonia on chest x-ray, headache, malaise), or conditions that may explain away respiratory symptoms and findings (e.g., congestive heart failure and musculoskeletal chest

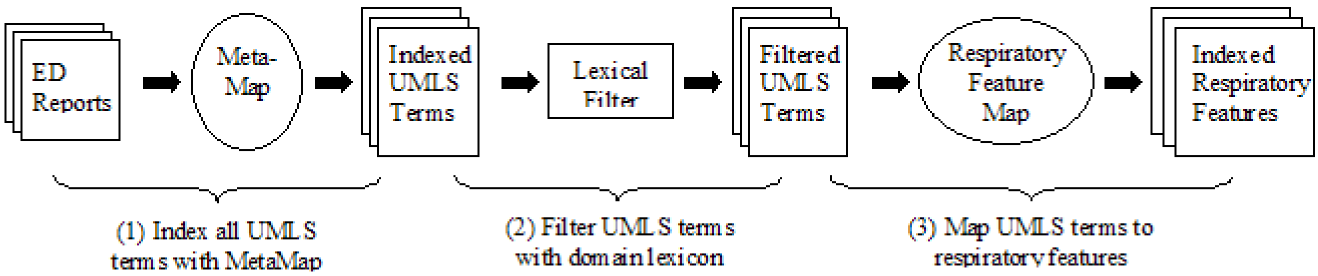

Figure 1 - To identify features that may help detect patients with a lower respiratory syndrome we indexed all UMLS concepts in a set of ED reports using MetaMap. MetaMap's output was filtered through a lexicon specialized for this domain, leaving only relevant UMLS concepts. Finally the indexed UMLS concepts were mapped to the 71 respiratory features

wall pain). The list can be viewed at http://omega.cbmi.upmc.edu/~chapman/respiratory-features.html.

Using MetaMap to Index Respiratory Features The process illustrated in Figure 1 was refined by manual review of MetaMap's output on a training set comprised of 50 visits to the University of Pittsburgh Medical Center (UPMC) ED during 2002. The training set was randomly selected from patients with a respiratory-related ICD-9 discharge diagnosis.

<u>*Step 1: Index all UMLS concepts with MetaMap*</u> We used all but one of the default settings for MetaMap, including selecting only the best concept and preferring precoordinated concepts. The default setting to only index simple noun phrases was changed to allow complex noun phrases with a prepositional phrase beginning with "of" to capture phrases like "shortness of breath" and "production of sputum."

If MetaMap did not generate at least one UMLS concept in the domain lexicon for any given phrase, the phrase was processed a second time – this time preferring multiple concepts. The granularity of the respiratory features does not always correspond directly to the granularity of UMLS concepts indexed from the Metathesaurus. For example, MetaMap's default setting to prefer single concepts will prefer the more specific UMLS concept "left sided chest pain (C0541828)" over the concept "chest pain (C0008031)". Because we do not distinguish in our domain lexicon between right- and left-sided chest pain, we only included "chest pain" in the lexicon.

Other solutions to this problem exist. One solution is to develop a more complete lexicon that contains parents and children of the relevant UMLS concepts or to implement rules based on the parent-child relationships in the Metathesaurus; however, compiling a complete lexicon for any domain would be expensive in terms of time and expertise, and using parent-child relationships may introduce errors. We implemented a simpler – albeit less graceful – solution of processing an unmapped phrase again for multiple concepts in case the less specific concept, which can be indexed to the head of the phrase (i.e., "chest pain" in our example), exists in the lexicon.

<u>*Step 2: Filter indexed UMLS concepts with a domain lexicon*</u> Tringali, et al. [6] showed that precision of indexing with MetaMap increases with a domain lexicon. We compiled a domain lexicon of UMLS concepts that map to the respiratory features as follows. First, we extracted a superset of cardiopulmonary findings and anatomy from the UMLS Metathesaurus by manually identifying three root concepts in the Metathesaurus and automatically extracting all of their children. Second, we used the interactive version of MetaMap to manually find UMLS concepts that matched the respiratory features. We included some UMLS concepts that occurred outside the superset (e.g., for headache and malaise).

Most respiratory features directly mapped to at least one UMLS concept (69/71). For example, the feature Wheezing maps directly to the UMLS concept "wheezing (C0043144)", and Chest Pain can map to any one of a list of UMLS concepts, including "chest pain (C0003031)", "angina pectoris (C0002962)", and "chest discomfort (C0235710)".

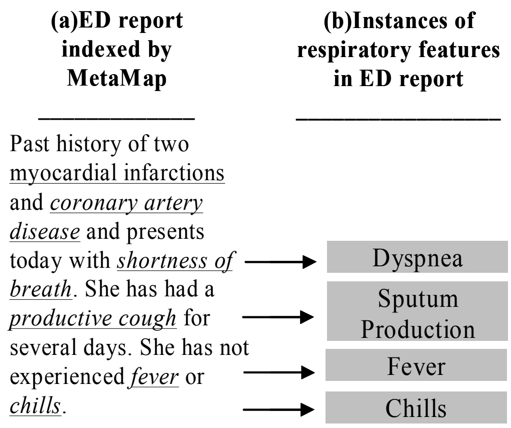

Figure 2. Target output of indexing. (a) A sample ED report after being indexed by MetaMap with UMLS concepts underlined and UMLS concepts in the domain lexicon also italicized. (b) Target output is a list of individual instances of acute respiratory features described in the report as occurring at the current ED visit. Four respiratory features should be annotated in this report. The UMLS term for "myocardial infarction" is not considered, because it is not in the domain lexicon. Because "coronary artery disease" occurred in the patient's past history the corresponding UMLS concept should not be mapped to the respiratory feature Coronary Artery Disease.

We also included in the lexicon UMLS concepts that may be combined to indirectly map to respiratory features for several reasons. First, two features had no direct map in the Metathesaurus but could be constructed indirectly from two atomic UMLS concepts. For instance, Poor Inspiration could be constructed with "breathing (C0004048)" and "poor – grade value

(C0542537)". Second, if MetaMap indexes a phrase with a more specific concept than exists in our lexicon, processing the phrase for multiple concepts may provide two UMLS concepts that in combination can map to a respiratory feature. Third, as described in Sneiderman, et al. [7], some clinical observations must be correlated with a qualitative or quantitative value to be considered an actual finding (e.g., "oxygen saturation of 99%").

Step 3: Map UMLS concepts to respiratory features Our algorithm checked first for direct maps within a phrase then for indirect maps. Indirect maps were comprised of either two UMLS concepts (UMLS-UMLS) or a UMLS concept and a numeric value (UMLS-numeric). To avoid generating false positives due to complex sentences, UMLS-UMLS combinations were restricted to concepts indexed within the same phrase. UMLS-numeric combinations, however, were allowed within the same sentence as long as the first number following or preceding the UMLS concept fell within a range of numeric values expected for that feature. For example, Fever would be indexed in the sentence "the patient's temperature was 38.5," because the UMLS concept "body temperature (C0005903)" was indexed, and the numeric value 38.5 fell between 38.0 and 44.0 or 101.5 and 113.

Evaluation We measured our ability to index the 71 respiratory features on a test set of 15 randomly selected patient visits to the UPMC ED during 2002. Inclusion criteria were a respiratory ICD9 discharge diagnosis and at least one ED report. We compared the automatically indexed respiratory features against respiratory features manually annotated by a physician board-certified in internal medicine and infectious diseases (JND) who used the manual annotation interface in GATE – a development environment for creating language engineering applications. GATE is available from the University of Sheffield (http://gate.ac.uk/) under the terms of the GNU General Public License. Any automatically annotated feature that overlapped with the manually annotated feature was counted as a true positive. We calculated the recall (sensitivity) and precision (positive predictive value) for the automatic indexing process with direct mapping only and with direct and indirect mapping. We performed a complete error analysis of the false negative and false positive maps to define the types of additional processing needed to successfully apply MetaMap to clinical reports.

Results

The 15 patient visits in the test set produced 28 separate ED reports. The physician annotator indexed 359 respiratory features in the 28 reports. Thirty-five of the 71 respiratory features occurred in the test set. The most frequently annotated respiratory features are shown in Table 1. The automatic indexing method performed with a recall of 0.55 (198/359) and a precision of 0.50 (198/399) when only mapping directly to UMLS concepts. When also allowed to map indirectly using atomic UMLS concepts, the recall increased to 0.72 (259/359) and the precision to 0.56 (259/460). Indirect mapping identified an additional 61 true positives and did not generate any false positives. Table 2 shows the distribution of false negative and false positive identification of respiratory features in the test set.

Discussion

Performance of the indexing process we applied was fairly good considering we basically used MetaMap "out-of-the-box" on a clinical indexing task requiring more knowledge than merely what UMLS concepts exist in the text. Results from our error analysis, described below, will potentially increase both recall and precision of the indexing process.

Table 1: Manually annotated Respiratory features in Test set with Frequency > 9

Respiratory features	Frequency
Fever	41
Rales/Crackles	26
Pneumonia Xray	25
Dyspnea	24
Chest Pain	22
Tachycardia	21
Cough	19
Tachypnea	19
Wheezing	16
Oxygen Desaturation	14
Sputum	14
Sweats	12
Chills	11
Cyanosis	11
Chest Tenderness	10

Error Analysis Errors fell into four broad categories, including problems with the domain lexicon, MetaMap errors, complications from manual annotation, and the need for contextual discrimination, which we discuss below.

Manual Annotation Over 60% of the false positives and 13% of the false negatives were related to the reference standard. Our category titled "possible annotation error" in Table 2 is subjective and could be challenged by another physician annotator. However, manual annotation is an imperfect process resulting from a tedious task laden with questions such as which terms should be included in the annotation (e.g., "severe chest pain" or "chest pain"), whether an uncertain diagnosis should be annotated (e.g., "I doubt the possibility of a pulmonary embolism"), and which concepts match the definitions requested by the researchers (e.g., in the sentence "the patient is complaining of fever times two days" is Fever a current problem for the patient). Errors due to manual annotation will always exist but can be reduced by generating a reference standard comprised of multiple physicians' annotations, as described by Hripcsak [8], and by implementing more complete and consistent annotator training with practice annotations on difficult or ambiguous cases.

Domain Lexicon Our lexicon was incomplete and generated 33% of the false negatives. In the test set we encountered terms we had not foreseen, such as "heart rate" in the context of tachycardia, "distant air sounds," and "submandibular lymphadenopathy." Some of the terms we had not seen in the training set map to UMLS concepts not included in the lexicon (e.g., "Heart Rate C0018810"), whereas some terms will need to be added as non-UMLS concepts (e.g., "submandibular"). Ten false negatives were due to not identifying terms that internally indicate the absence of the condition, such as "afebrile" or "nondiaphoretic."

A more troublesome dilemma in constructing the lexicon involves vague terms in the text that can only be interpreted with extensive contextual knowledge. For instance, the respiratory feature Chest Congestion is often expressed in the text with only the word "congestion". However, "congestion" can also mean nasal congestion, and the difference is not always clear from the immediate context.

Table 2: Etiology of Errors in Test Set

False Negatives (n = 100)	Freq.
Domain Lexicon (33%)	33
Internal negation of term	10
New lexical variant	13
Vague term in text	10
MetaMap Mistake (29%)	26
Phrasal syntax inadequate	16
Lexical variant not indexed	10
Manual Annotation (13%)	14
Non-overlapping boundaries	9
Possible annotation error	5
Need Contextual Discrimination (25%)	27
Implied information in text	20
Section identification required	7
False Positives (n = 201)	
MetaMap Mistake (1%)	2
Manual Annotation (62%)	124
Non-overlapping boundaries	2
Possible annotation error	95
Interpretation of current visit	14
Uncertainty in text	13
Need Contextual Discrimination (37%)	75
Implied information in text	4
Section identification required	60
Finding verification required	11

<u>MetaMap Mistakes</u> MetaMap produced only a tenth of the total errors in the test set, and all but two of the 28 errors were false negatives. Lexical variants not recognized by MetaMap include "diaphoretic" instead of "diaphoresis," "respires" instead of "respiration," "pO2" instead of "percent O2," and "rhonchourous" instead of "rhonchi." Sixteen errors were due to information about the respiratory feature extending across phrasal boundaries, as in "chest wall examination did demonstrate some slight tenderness when the patient had pressure applied to the right side of the thoracic cage."

<u>Contextual Discrimination</u> False positives were largely due to the need for what we call contextual discrimination within the report. *Finding validation* based on the context is necessary to avoid annotating "cough medicine" and "worsened by coughing" as instances of Cough. *Temporal discrimination* is necessary to determine whether the condition occurred in the past history, is a current problem, or is mentioned as a hypothetical possibility (e.g., "should return if fever develops"). *Anatomic location discrimination* is vital to discriminating among interpretations of ambiguous terms like "mass" that could indicate Pulmonary Mass or a non-pulmonary mass not included in our feature list. *Finding-disease discrimination* is important for clinical concepts like pneumonia, which could appear in several places on the finding-disease continuum and is in our respiratory feature list as a historical finding (Pneumonia History), a radiological finding (Pneumonia on Chest Radiograph), and a diagnosis (Pneumonia Diagnosis). *Contextual inference* would enable an automated system to make inferences a physician easily makes regarding features not explicitly mentioned in the text in sentences like "Chest x-ray was normal" or "Lung sounds were clear." In these sentences, respiratory features such as Pneumonia on Chest Radiograph, Pneumothorax, Rales/Crackles, and Wheezing can be annotated even though they were not explicitly mentioned.

Many of the false positives due to contextual discrimination can be eliminated with identification of the report section in which the feature occurs. Report sections may be full paragraphs delineated by a heading (e.g., Past Medical History, Lungs, HEENT). More often, though, in ED reports the relevant section may only comprise a single sentence or part of a sentence, as in "The patient has a history of shortness of breath and presents today with chest pain." We are testing keyword-based algorithms for detecting the beginning (e.g., "history of") and the end (e.g., "presents") of historical, radiological, and hypothetical conditions.

Related Work Other researchers have applied MetaMap to clinical reports. Indexing arterial branching predicates in cardiac catheterization reports, Rindflesch [9] reported recall of 0.83 recall and precision of 1.0. Tringali et al. [6] indexed UMLS concepts in esophago-gastroduodenoscopy reports with a recall of 0.62 and a precision of 0.76. Both of these indexing tasks differed from ours in that we were not directly measuring MetaMap's ability to index UMLS concepts but were measuring our ability to index the UMLS concepts and then map them to an externally defined set of clinical concepts representing symptoms, findings, and diseases the patient exhibited at the hospital visit.

Identification of clinical concepts in patient reports has been the focus of research by Sager [10], Friedman [11], Haug [12], Baud [13], Hahn [14], Taira [15], and others who have developed medical language processing systems from the ground up. We wanted to determine how successfully we could apply a pre-existing, publicly available indexing technique to the task for rapid implementation in a biosurveillance system.

Limitations and Future work We will use the results of this study to enrich our domain lexicon and to guide our implementation of post-processing techniques. If we can increase recall and precision sufficiently (the meaning of sufficient performance is another paper in and of itself), we may implement the indexing process into the Real-time Outbreak and Disease Surveillance (RODS) system [16] for automatic detection of patients with respiratory illnesses such as SARS.

A major limitation of this study was a reference standard comprised of a single physician. We will use the test set in this pilot study to train multiple physicians for a future reference standard.

In this study we did not address the critical task of determining whether a feature is described as present or absent in a report. In the test set, 47% of the respiratory features were manually annotated as being absent. The future version of our indexing application will employ and evaluate a regular expression-based negation algorithm called NegEx [17].

Conclusion

The purpose of this project was to examine the ability of an indexing application created for biomedical texts to identify respiratory features described in ED reports. We demonstrated the usefulness of MetaMap's indexing techniques for this task. We believe our methods for compiling and implementing a domain lexicon of UMLS concepts are generalizable to other domains within and outside of biosurveillance. Regardless of the type of indexing technique used to index UMLS phrases in clinical reports, the error analysis we provided can be a useful road map for the types of processing necessary in order to successfully map UMLS concepts to clinical concepts.

Acknowledgments.

This research was conducted as part of the NLM Rotation Program at the Lister Hill National Center for Biomedical Communications at the National Library of Medicine, National Institutes of Health, Department of Health. The authors express thanks to Alexa McCray, Olivier Bodenreider, Alan Aronson, and May Cheh for their assistance and support. Partial support also provided by AHRQ Grant No. 290-00-0009.

References

[1] http://www.who.int/csr/don/2003_04_22/en/. Accessed April 23, 2003.

[2] Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, et al. A Major Outbreak of Severe Acute Respiratory Syndrome in Hong Kong. *N Engl J Med* 2003.

[3] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001:17-21.

[4] Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. *J Am Med Inform Assoc* 1998;5(1):1-11.

[5] McCray AT, Nelson SJ. The representation of meaning in the UMLS. *Methods Inf Med* 1995;34(1-2):193-201.

[6] Tringali, Michele; Marcelo Fiszman; Thomas Rindflesch; and Olivier Bodenreider. Automatic extraction of concepts from gastrointestinal endoscopy reports. *Medical Informatics, Europe* 2003 [abstract].

[7] Sneiderman CA, Rindflesch TC, Aronson AR. Finding the findings: identification of findings in medical literature using restricted natural language processing. *Proc AMIA Annu Fall Symp* 1996:239-43.

[8] Hripcsak G, Kuperman GJ, Friedman C, Heitjan DF. A reliability study for evaluating information extraction from radiology reports. *J Am Med Inform Assoc* 1999;6(2):143-50.

[9] Rindflesch TC, Bean CA, Sneiderman CA. Argument identification for arterial branch predications asserted in cardiac catheter reports. *Proc AMIA Annu Fall Symp* 2000:204-8.

[10] Sager N, Friedman C, Lyman M. *Medical language processing: computer management of narrative data.* Reading, Massachussets: Addison Wesley; 1987.

[11] Friedman C. A broad-coverage natural language processing system. *Proc AMIA Symp* 2000:270-4.

[12] Haug PJ, Koehler S, Lau LM, Wang P, Rocha R, Huff SM. Experience with a mixed semantic/syntactic parser. *Proc Annu Symp Comput Appl Med Care* 1995:284-8.

[13] Baud RH, Lovis C, Ruch P, Rassinoux AM. A light knowledge model for linguistic applications. *Proc AMIA Symp* 2001:37-41.

[14] Hahn U, Romacker M, Schulz S. MEDSYNDIKATE-a natural language system for the extraction of medical information from findings reports. *Int J Med Inf* 2002;67(1-3):63-74.

[15] Taira RK, Soderland SG. A statistical natural language processor for medical reports. *Proc AMIA Symp* 1999:970-4.

[16] Tsui FC, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. *J Am Med Inform Assoc* 2003:(in press).

[17] Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform* 2001;34(5):301-10.

Address for correspondence

Wendy W Chapman,

Center for Biomedical Informatics,

Suite 8084 Forbes Tower, 200 Lothrop Street, Pittsburgh, PA 15213. 412-383-8140;

chapman@cbmi.pitt.edu