Natural Language Processing for a Pediatric Cardiology Problem List

Ashley Phimister, Lee A. Pyles, MD. University of Minnesota Amplatz Children's Hospital, Minneapolis, MN

Abstract:

Background/Hypothesis: Natural Language Processing (NLP) is a data analysis technique that evaluates of the meaning of text passages by a computer. NLP programs perform a step-wise analysis of the text to assign functionality to words then phrases and sentences and finally entire documents. Using NLP to parse problem lists and medication lists from discharge summaries for pediatric congenital heart disease patients will provide an efficient way to create a summary of a health care record derived from health care notes. We hypothesize that a principal diagnosis can be created from a discharge summary with NLP using a standard glossary of pediatric cardiac terms such as that of the Society of Thoracic Surgeons (STS) database or the Pediatric Cardiac Care Consortium (PCCC).

Methods: We used our University of Minnesota BioMedical Information Collection and Understanding System (BioMedICUS) NLP computer application to find pediatric cardiac principal diagnoses. The parser was alternatively augmented either with the STS or PCCC set of terms and with a supplementary list of cardiac cognates, abbreviations and acronyms. The parser outputs were then analyzed by hand by the medical student.

Results: The NLP method using the PCCC and STS term list provided sensitivity for the primary cardiac diagnosis with the first weighted phrase of 82% +/- 0.85%SE and 77% +/- 0.97% respectively (P=NS, n=302). For the PCCC glossary the first 5 highest-weighted phrases contained the primary cardiac diagnosis 96% of the time with the average line in which the primary cardiac diagnosis was found being 1.8 +/- 3.1. There were 2 times that this method was unable to identify the primary cardiac diagnosis in any line. Using the STS glossary, the first 5 highest-weighted phrases contained the primary cardiac diagnosis 88% of the time with the average line in which the primary cardiac diagnosis was found being 1.9 +/- 2.9. There were 19 times that this method was unable to identify the primary cardiac diagnosis in any line.

Conclusion: We are able to devise an NLP protocol using the BioMedICUS application that can identify a sentence containing a principal cardiac diagnosis. This early use of NLP in the pediatric cardiovascular domain offers promise to facilitate EHR implementation for these patients, including problem list, ordering, and emergency summary functions.

Background:

Natural Language Processing (NLP) is a data analysis technique that evaluates the meaning of text passages by a computer. NLP programs perform a step-wise analysis of the text to assign functionality to words then phrases and sentences and finally entire documents. Using NLP to parse problem lists and medication lists from discharge summaries for pediatric congenital heart disease patients will provide an efficient way to create a summary of a health care record derived from health care notes. It will reduce the input needed from a health care provider into the summary, while providing a concise summary of a pediatric cardiac patient status that can be accessed quickly. In addition, a precise summary will help to reduce errors such as repetition of tests, unnecessary hospitalizations, and avoidable poor outcomes.

Hypothesis:

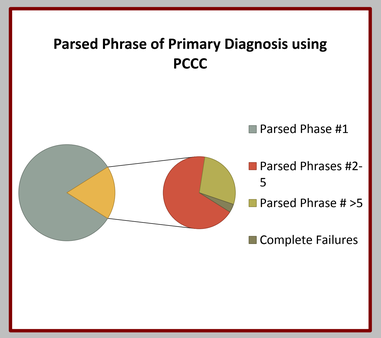
We hypothesize that a principal diagnosis can be created from a discharge summary with NLP using a standard glossary of pediatric cardiac terms such as that of the Society of Thoracic Surgeons (STS) database or the Pediatric Cardiac Care Consortium (PCCC). The goal of this project is to create a concise summary of pediatric cardiac patient status that can be ported to a personal health record or an EHR, thus limiting the need for health care provider input. Obtaining provider input into initiation of a personal health record (PHR) has been a major barrier to successful implementation of accurate PHRs.

Methods:

- BioMedICUS has been developed at the University of Minnesota in the NLP Center of the Institute for Health Informatics. BioMedICUS is based on an industrial strength Unstructured Information Management Architecture (UIMA) and consists of adaptable machine learning and rule-based components designed to address the specifics of the EHR implementations at participating clinical sites.
- The BioMedICUS application was used initially with the Unified Medical Lexicon System to identify Content Unique Identifiers (CUIs) specific to pediatric cardiovascular disease as our first attempt to identify the principal cardiac diagnosis. This failed to identify important diagnostic phrases.
- BioMedICUS also includes a Stanford Parser algorithm that employs weighted phrase analysis to identify the cardiac terms. The Pediatric Cardiac Care Consortium and STS terminology lists (lexicons) were augmented with cognates, abbreviations and acronyms to aid analysis.
- The result of the weighted phrase analysis was hand-inspected by the medical student after learning about congenital heart disease in the clinical setting with supervision from the pediatric cardiology mentor.
- Evaluation of results consisted of inspection of the BioMedICUS output in order of term frequency to determine the first phrase that included the principal cardiac diagnosis.

Table 1. Results

	PCCC	STS (IPCCC)
Primary Diagnosis in First Line (n = 300)	246	230
Sensitivity	82% ± 0.85% SE	77% ± 0.97% SE (p = NS)
Average Line of Primary Diagnosis	1.8	1.9
Standard Deviation of Line of Primary Diagnosis	3.1	2.9
Number of complete Failures to identify any diagnostic phrase	2	19
Primary Diagnosis within First 5 Phrases	96%	88% (p = 0.0003)



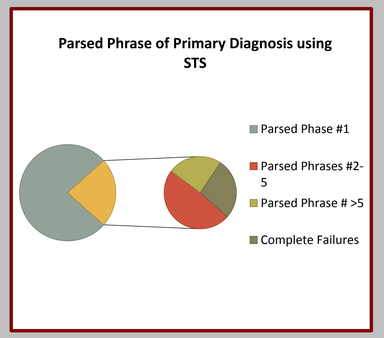


Table 2. Number of Successes and Failures by Diagnosis

Diagnosis	Total Number of Diagnosis	Number of Successes (Diagnosis in Parsed Phrase #1) for PCCC	Number of Complete Failures for PCCC	Number of Successes (Diagnosis in Parsed Phrase #1) for STS	Number of Complete Failures for STS
Aortic Insufficiency	1	1		1	
Aortopulmonary Window	1	1			
Aortic Stenosis	8	5	1	6	2
Atrial Septal Defect	28	24		25	
Atrioventricular Canal	20	15		17	
Bicuspid Aortic Valve	1	1		1	
Cardiac Rhabdomyoma	1			1	
Cardiomyopathy	5	5		4	
Coarctation of the Aorta	23	12	1	17	1
D Transposition of the Great Arteries	31	24		14	4
Double Inlet Left Ventricle	2	2		2	
Double Outlet Right Ventricle	10	9		9	
Double Aortic Arch	2	2		2	
Ebstein's Anomaly	2	2		2	
Heart Block	2	2		2	
Heterotaxy	5	4		5	
Hypoplastic Left Heart	14	14		14	
Hypoplastic Right Heart	1	1		1	
Hypoplastic Right Ventricle	1	1		1	
Interrupted Aortic Arch	7	7		7	
L Transposition of the Great Arteries	3	3		3	
Left Atrial Myxoma	1				
Malposed Great Vessels	1	1		1	
Mitral Stenosis	1	1		1	
Myocarditis	1	1			1
Obstructed Left Heart	1			1	
Partial Anomalous Pulmonary Venous Return	5	5		5	
Patent Ductus Arteriosus	6	2		2	
Patent Foramen Ovale	3	2		2	
Pulmonary Atresia	4	4		4	
Pulmonary Stenosis	4	3		4	
Right Pulmonary Artery Anomaly	1	1		1	
Single Atrium	1	1		1	
Single Ventricle	2	2		2	
Subaortic Membrane	3	3			3
Subaortic Tunnel	1	1		1	
SVC Syndrome	1	1		1	
Total Anomalous Pulmonary Venous Return	2	2		2	
Tetralogy of Fallot	39	34		21	8
Tricuspid Atresia	3	3		2	
Truncus Arteriosus	5	3		5	
Ventricular Septal Defect	47	41		40	
Total	**300**	**246**	**2**	**230**	**19**

Diagnosis in Different Lines for PCCC and STS:

Diagnosis: D-Transposition
PCCC:
Parsed Phrase #1: *"a 5-week-old male status post arterial switch for de-transposition of the great arteries on February 11 , 2005 , who presents with poor feeding and fussiness",1,6*
STS:
Parsed Phrase #1: *"Normal S1 and normal S2 , 2/6 systolic ejection murmur at the left lower sternal border .",1,3*
Parsed Phrase #2: *"a 5-week-old male status post arterial switch for de-transposition of the great arteries on February 11 , 2005 , who presents with poor feeding and fussiness",1,3*

Diagnosis in Line 1 for Both PCCC and STS

Diagnosis: AV Canal
PCCC:
Parsed Phrase #1: showed a partial AV canal with a small primum atrial septal defect with left to right shunt,1,11
STS:
Parsed Phrase #1: showed a partial AV canal with a small primum atrial septal defect with left to right shunt,1,10

Example of Success with PCCC and Complete Failure with STS

Diagnosis: Tetralogy of Fallot
PCCC:
Parsed Phrase #1: "a 7-month-old infant with tetralogy of Fallot who came in on the on 03/12/2007 and underwent a full repair his tetralogy of Fallot",1,5
STS: Complete failure – unable to find diagnosis
Parsed Phrase #1: "was discharged home on postoperative day # 4 in the care of his parents on the following medications , Lasix 5 mg p.o.",1,1

Example of Failure using STS:

Diagnosis: Aortic Stenosis
Discharge Summary:
Discharge Summary-FUMC documenttext= FINAL
ADMIT DIAGNOSES:
1. Aortic stenosis status post balloon aortic valvotomy in 08/1996.
2. Status post Ross procedure in 1998 with right ventricular outflow tract homograft conduit.
3. Status post right ventricular outflow tract conduit replacement with a 25-mm homograft and a new aortic root reconstruction in 08/2006.
4. Patient with significant right ventricular outflow tract obstruction.

DISCHARGE DIAGNOSES:
1. Aortic stenosis status post balloon aortic valvotomy in 08/1996.
2. Status post Ross procedure in 1998 with right ventricular outflow tract homograft conduit.
3. Status post right ventricular outflow tract conduit replacement with a 25-mm homograft and a new aortic root reconstruction in 08/2006.
4. The patient with significant right ventricular outflow tract obstruction, status post placement of a conduit stent with improved gradient across the right ventricular outflow tract.

STS Parsing:
MRN,Phrase,Frequency,Diag Words
_1,post right ventricular outflow tract conduit replacement with a 25-mm homograft and a new aortic root reconstruction in 08\/2006,2,6
_1,"The patient with significant right ventricular outflow tract obstruction , status post placement of a conduit",1,6
_1,"significant right ventricular outflow tract obstruction , status post placement of a conduit",1,6
_1,"with significant right ventricular outflow tract obstruction , status post placement of a conduit",1,6
_1,"significant right ventricular outflow tract obstruction , status post placement",1,5
_1,right ventricular outflow tract conduit replacement with a 25-mm homograft and a new aortic root reconstruction,2,5
_1,right ventricular outflow tract conduit replacement,2,4
_1,post Ross procedure in 1998 with right ventricular outflow tract homograft conduit,2,4
_1,stenosis status post balloon aortic valvotomy in 08/1996, 2, 4
_1,significant right ventricular outflow tract obstruction,1,3

Conclusion:

We are able to identify a phrase containing a principal cardiac diagnosis using an NLP protocol with a Stanford parser algorithm in the BioMedICUS application. This early use of NLP in the pediatric cardiovascular domain offers promise to facilitate EHR implementation for these patients. Ability to identify a principal diagnosis will facilitate activities such as creation of a problem list, ordering, and emergency summary functions. The PCCC list of terms may be more useful for this sort of Natural Language Processing analysis as opposed to the STS lexicon due to the tree-structure listing of those terms.

Acknowledgements: the authors acknowledge the support of the American Heart Association Helen N. and Harold B. Shapira Heart Medical Scholarship for Ms. Phimister, the assistance of Serguei Pakhomov, PhD, Genevieve Melton-Meaux, MD and Robert Bill of the UM Institute for Health Informatics and Pouya Hemmati, MS III, and support for Dr. Pyles from University of Minnesota Interdisciplinary Informatics Research Grant and University of Minnesota Healthy Informatics ARRA Training Grant #1T15OC000073-01. Dr. Pyles was a University Partnership for Health Informatics Scholar in the UM Health Informatics MS Program.