

## **Barriers to NLP Adoption**

- We have a long history of resarch on NLP methods in the clinical domain [1].
- However, the complexity of unstructured clinical text makes analysis a hard problem and its accuracy varies.
- Domain experts may be able to fix problems with the models but they may not be familiar with symbolic and machine learning techniques.

## **Design Requirements**

We have built upon ideas in Visualization, Interactive Machine Learning and Interface Design research.

Our design requirements are summarized as follows:

- R1: The tool should make it easier for machine learning non-experts to work with NLP models.
- R2: It should incorporate efficient mechanisms for annotation and labeling, and also for encourage feedback that is consistent and informative.
- R3: The interactive components should support the entire interactive machine learning loop - i.e. a review, feedback and retrain cycle.

## Acknowledgments

Our demo uses an example dataset of colonoscopy reports and is based on the work done by Harkema et. al. [2]. This research is supported by NIH grant 5R01LM010964.

### References

- 1 Chapman, W. W., Nadkarni, P. M., Hirschman, L., D'Avolio, L. W., Savova, G. K., and Uzuner, O. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. Journal of the American Medical Informatics Association 18, 5 (2011), 540–543.
- [2] Harkema, H., Chapman, W. W., Saul, M., Dellon, E. S., Schoen, R. E., and Mehrotra, A. Developing a natural language processing application for measuring the quality of colonoscopy procedures. JAMIA 18, Supplement (2011), 150–156.
- [3] Wattenberg, M., and Viegas, F. B. The word tree, an interactive visual concordance. *IEEE Transac*tions on Visualization and Computer Graphics 14, 6 (2008), 1221-1228.
- [4] Brooke, J. SUS: a quick and dirty usability scale. In Usability evaluation in industry, P. W. Jordan, B. Weerdmeester, A. Thomas, and I. L. McLelland, Eds. Taylor and Francis, London, 1996.

# **Bridging the Natural Language Processing Gap: An Interactive Clinical Text Review Tool**

Gaurav Trivedi<sup>1</sup>, Phuong Pham<sup>1</sup>, Wendy Chapman<sup>2</sup>, Rebecca Hwa<sup>1</sup>, Janyce Wiebe<sup>1</sup>, Harry Hochheiser<sup>1</sup> <sup>1</sup>University of Pittsburgh, Pittsburgh, PA; <sup>2</sup>University of Utah, Salt Lake City, UT trivedigaurav@pitt.edu

# **Interface Design**

EMR	VisW	eb		ASET		model	EL ¥ List.0.	l	any-a	adeno	ma Ŧ	j is ma	irkea	as taise	e for this record. Label it: True
doc #	any-adeno	appendicea	858	biopsy	cecum	ileo-cecal-v	indication-ty	informed-	nursing-rep	no-prep-ade	not-prep-ad	yes-prep-ad	proc-aborte	withdraw-ti	Document View
		ea				-5	-j	8	ę	d	ä	å	rte	É.	Report: #0001
0005	alse	True	False	False	?	True	False	True	False	False	False	True	?	?	E
0022	True	True	False	True	?	True	False	True	False	False	False	True	?	?	<b>U</b>
0027	False	True	False	False	?	True	False	True	False	False	False	True	?	?	0_H
0033	False	True	False	False	?	True	False	True	False	False	False	True	?	?	Report ID Pat
0038	True	True	False	True	?	True	False	True	False	False	True	False	?	?	1,lGOCGBrZJw6P lGO
0050	True	True	False	True	?	True	False	True	False	False	False	False	?	?	E_0_H
0053	True	True	False	False	?	True	False	True	False	False	True	False	?	?	[Report de-identifi
0057	True	True	False	True	?	True	False	True	False	False	False	True	?	?	
0065	True	True	False	False	?	True	False	False	False	False	False	True	?	?	COLONOSCOPY REPORT
0070	False	True	False	False	?	True	False	True	False	False	False	True	?	?	COLONOSCOPT REPORT
0074	True	True	False	True	?	True	False	True	False	False	False	True	?	?	ID#: **ID-NUM PRO
0097	True	True	False	True	?	True	False	True	False	False	True	False	?	?	NAME: **NAME[AAA,
rue ind 0.2 0.4 2 care fragme	nt		a		80 60 40 20		False	<b>B</b>		Tr a f r a	isk erform Iternati	cators ality ve	D		ASSISTANT: ***NAM DOB: 05/17/1951 W AGE: 56 DATE: 0 ASA CLASS: Class I INDICATIONS: 1) ch MEDICATION: 200 mi Benadryl IV PREPARATION OF COLO
False ind scope advanc note history patient	e	•								s r a	alse ind cope ormal dvance oreparat oatient	)			FINDINGS: After the thoroughly explained <b>performed</b> and revea introduced through of the prep was good was fully examined.

A The Grid view shows the extracted variables in column overview of NLP results. Below the grid, we have statistics a of the classifications and  $\bigcirc$  the list of top indicators aggreg D Indicators from the active report are shown on the right. the patient reports with the indicator terms highlighted. bar on the top, or by using the right-click context menu.

The WordTree [3] view provides the ability to search | The for and explore word sequence patterns found across cludi the documents in the corpus, and to provide feedback chan to retrain NLP models.

		"hot biopsy" indicates	biopsy 🗘	to be: True   False		
	others were			(		Documer
	raphed and	ed with hot		forceps-		Feedba
	as removed		•	loroopo	•	1. "prep
	num pieces	by by	iopsy {		— taken — .	
	num pieces	cm		was	- performed using the erbe .	<b>2.</b> #0022
	e , initials cmk and polyp s	sigmoid		2 much of the colon	had thick liquid stool obscuring the mucosa .	n Clea
5	S	a		using the erbe .		

A demo video of the tool is available at http://vimeo.com/trivedigaurav/emr-demo.

# **Our Solution: An Interactive Tool for NLP on Clinical Text**

Our goal is to close the NLP gap by providing clinical researchers with highly-usable tools that will facilitate the process of reviewing NLP output, identifying errors in model prediction, and providing feedback that can be used to retrain or extend models to make them more effective. We have developed an interactive web-based tool that facilitates both the review of binary variables extracted from clinical records, and the provision of feedback that can be used to improve the accuracy of NLP models.

	User
False F	We co cians
ordTree View 📿 Re-Train	insight
Variable any-adenoma +	associa opport
De-ID copyright (c) 1999-2009 University of Pittsburgh. All rights reserved.S_	menta
ient ID Principal Date Record Type CGBrZJw6P 20080327 1003 OP	We us
ed (Limited dataset compliant) by De-ID v.6.22.08.0]	of 10-
	a glob
CEDURE: Colonoscopy, diagnostic	The a
BBB] *≈*NAME[CCC: YYY ZZZ], MD V:	A sun
E[XXX WWW], RN and ***NAME[VVV UUU], RN ARD: outpatient 3/27/2008	user st
[	
ange in bowel habits crograms Fentanyl IV, <mark>6</mark> milligrams Versed IV, 50 milligrams	Cate
1: good	Work
<b>risks</b> benefits and <b>alternatives</b> of the procedure were d, informed consent was obtained. Digital rectal exam was led no <b>abnormalities</b> . The Pentax EC-3870LK endoscope was the anus and <b>advanced</b> to the ascending colon. The quality d. The instrument was then slowly withdrawn as the colon It was not possible to reach the cecum.	
ns and individual documents in rows, providing an	
about the active variable with $\bigcirc$ the distribution	L
gated across all the documents in the dataset.	
. $\bigcirc$ The <b>Document view</b> shows the full-text of	e Word
Feedback can be sent using the yellow control	
<b>Re-Train view</b> lists user-provided feedback, in- ing any potential inconsistencies, and specifies	
nges in variable assignments due to retraining.	Feedb
ent View WordTree View C Re-Train 2	_
Mark all documents as unread   Remove bolded change markers	Re-Tr
was good" indicates <b>yes-prep-adequate</b> to be <b>true</b> .	×
2 should have <b>any-adenoma</b> marked as <b>false</b> .	×
ear All C Re-Train	
	Future



# **Study and Results**

onducted a formative user study with five cliniand clinical researchers as participants to gain it into usability factors of the tool that may be lated with errors or confusion, and to identify tunities for improvement via re-design or impleation of new functionality.

sed the **System Usability Scale** [4] consisting questions on a 5-point Likert scale to help get cal view of subjective assessments of usability. verage SUS score was 70.5 out of 100.

mary of recommendations inferred from the tudy for is given below:

egory	Recommendation
cflow	1. Allow sorting (or filtering) of the docu- ments in the grid based on the prediction probabilities. This would make it easier for the users to prioritize documents to review.
	2. Add a button to open the next-in-line doc- ument for review. The order may be de- cided either trivially based on ID num- ber or by using an active learning ap- proach. This would save the users to navi- gate through the grid when they don't have their own strategy for selecting documents for review.
lTree	1. Change the layout of the tool to show the WordTree view along with the document view. This would allow the user to quickly go through the full report text when the wordtree tree is unable to provide sufficient contextual information.
	2. Allow selection of multiple branches in the tree to give feedback on multiple paths in the tree at once.
back	1. Provide a feedback mechanism to specify that a text span does not indicate either of the classes. This would allow the user to remove non-informative but possibly mis- leading features in re-training.
raining	1. Perform auto-retraining in the background when a sufficient number of feedback items have been provided by the user.
	2. Provide a built-in mechanism to validate and generate a performance report for the current model against a held-out test set.

e efforts will involve incorporating these recommendations and conducting an empirical evaluation.