Peeling Away the Black Box Label: Clinical Validation of a MaxEnt Machine Learning Character N-gram Feature Set for Acute Lung Injury

Imre Solti, MD, PhD[1], Colin R. Cooke, MD, MSc[2], Fei Xia, PhD[3], Mark M. Wurfel, MD, PhD[4]

[1] Biomedical and Health Informatics, [2,4]Pulmonary and Critical Care Medicine, [3]Linguistics; [1,3,4]University of Washington, Seattle, WA; [2]University of Michigan, Ann Arbor, MI

Abstract

Peeling away the "black box" label from machine learning approaches will increase trust and acceptability for machine learning among clinicians. In our presentation we show an example of a natural language processing and MaxEnt-based document classification algorithm. Top ranked character n-gram features for Acute Lung Injury were analyzed and validated using clinical expertise. These results provide an example of the intuitive link between clinical expertise and machine learning approaches that will help to advance the acceptance of these approaches by clinicians .

Introduction

Machine learning approaches are often labeled “black boxes” due to their lack of transparency. While this label is true for some machine learning algorithms, others provide ready clues that make explaining the results easier. Studies on intelligent systems show that explanation of results increase the trust and acceptability of machine learning approaches.[1]

As part of a clinical Natural Language Processing (NLP) project we developed a machine learning-based radiology report classification system for Acute Lung Injury (ALI) patients. We present the performance results of the machine learning based system in other venue.[2] In this presentation we will show how we used clinical expertise to validate the top ranked character n-gram features of a MaxEnt machine learning algorithm.

Data and Methods

953 chest x-ray reports were manually classified by clinicians specializing in Pulmonary and Critical Care Medicine as ALI positive or negative. These domain experts generated a list of 48 keyword phrases that they considered specific indicator of the patients' ALI status. We processed the radiology reports with the MALLET machine learning package using the MaxEnt algorithm.[3] We trained MaxEnt on the character n-grams as features.[2]

MALLET generates models as binary objects. We used the package's wrapper to convert the model to text. After sorting the features by weight we analyzed the features with highest positive weight (above a threshold) and checked how well the character n-grams overlap with the keyword phrases that were provided by the experts.

Results

A subset of the results is shown below (Table 1) for selected top ranked character six-grams and their overlap with the phrases on the experts' list.

N-gram Feature	Keyword Phrase
_opaci	**_opaci**ties_
y_opac	_patch**y_opac**ities_
teral_	_bila**teral_**
a_and_	_edem**a_and_**

Table 1. Sample six-gram features and keywords

Conclusion

Our study shows that for some combined NLP and machine learning algorithms the visualization of the results can provide a window on these procedures that will be readily appreciated by clinicians. We suggest that providing such intuitive examples of human-computer interactions in clinical informatics will improve the acceptance of NLP and machine learning approaches by clinicians.

References

1. Lim BY, Anind KD, Avrahami D. Why and why not explanations improve the intelligibility of context-aware intelligent systems. Computer Human Interaction 2009, Boston MA, April 4-9, 2009.
2. Solti I, Cooke CR, Xia F, Wurfel MM. Automated classification of radiology reports for acute lung injury: Comparison of keyword and machine learning based natural language processing approaches. NLP Workshop, IEEE International Conference on Bioinformatics and Biomedicine, 2009 (Accepted for publication).
3. MALLET: Toolkit. http://mallet.cs.umass.edu/ Accessed at September 24, 2009.