Classifying Physician Needed Response Questions on Online Health Forums

Team Emma:
Kristen McGarry
Deahan Yu

Background

In a 2001 research study, **~40%** of respondents with Internet access reported **using the Internet to look for health/health care advice**, with ⅓ of those respondents reporting that information found **affected a health/healthcare decision**.[1]

Online health forums (also called web medical forums) provide a **knowledge-sharing platform** for discussions between patients, patients loved ones/caregivers, and physicians for medical consultations.[2]

Patient communities serve as a **social network**[2] & **support mechanism** for patients and loved ones to receive **informational and emotional support**.[3]

The **natural-language content** allows the users to fully express their information needs without knowing medical lingo.[4]

Problem

With ⅓ of respondents reporting that using the Internet **affected a health/healthcare decision**[1], it is vital that those individuals do not receive potentially **harmful, incorrect, incomplete or bias information**.[5]

In general, **physicians feel obliged** to provide online consultations, but feel the new responsibility comes with **increased burden** (i.e. time, responsibility, lack of background knowledge on a given case).[6]

In a 2015 study (n=6,880), **54.4%** (3,680) of physicians reported at least 1 symptom of **burnout**, representing a **10% increase** since 2011 (45.5%,3,310).[7]

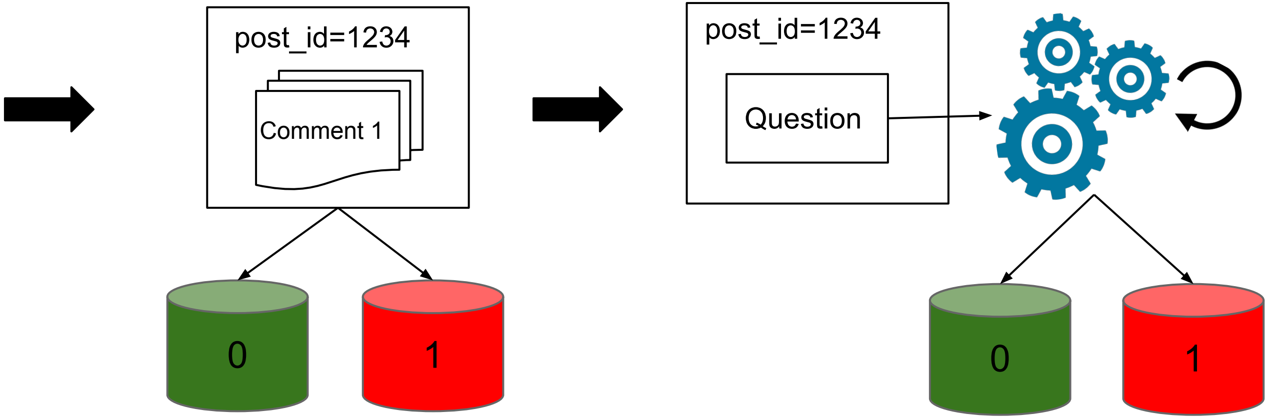
With physician burnout at an all time high, and with increased demand for physician responses on online health forums, how do patients receive reliable medical advice online?

Methods

Data:
1,403,220 questions
2,366,387 follow-up comments

Task 1:
Question Labeling Task based on Comments

Task 2:
Question Classification Task

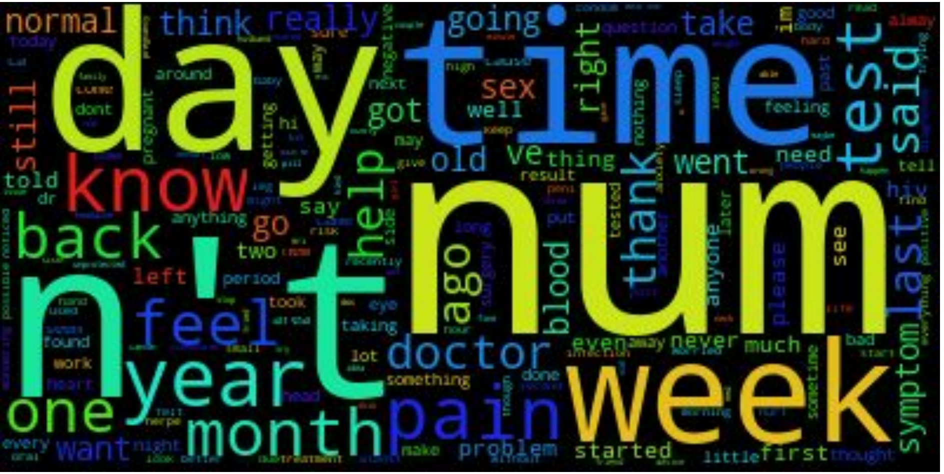


Task 1: Question Labeling based on Comments

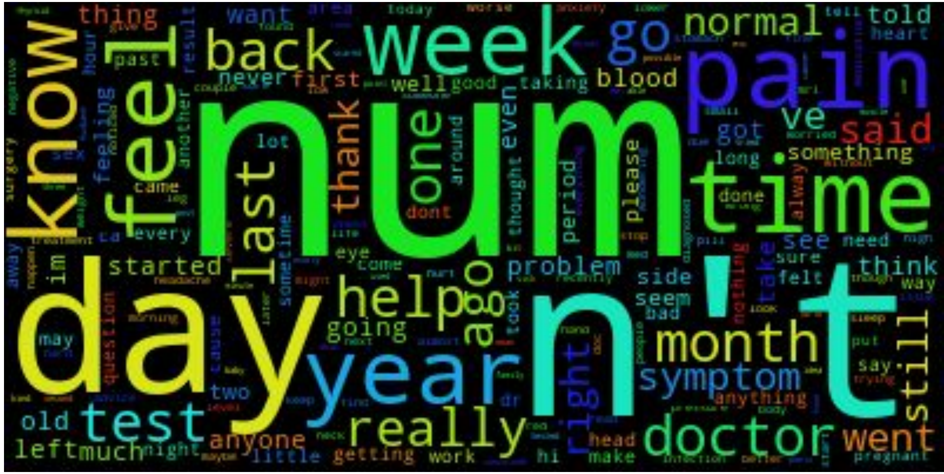
Initial Regex	Random 200 comments		
Labeled 1’s and 0’s	Cohen’s Kappa = 0.73		
Patterns & Errors	Finalized Regex		
	Prec	Recall	F1
	0.76	0.91	0.83
Labeled comments → questions	~23% labeled as 1’s		

Task 2 Question Classification

- Pre-processing
  - removed stop-words
  - replaced numeric values with “num”
- Word frequency analysis



Label = 0

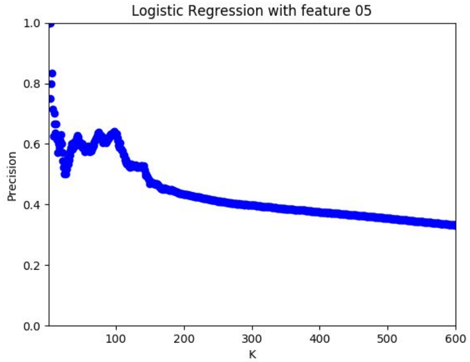


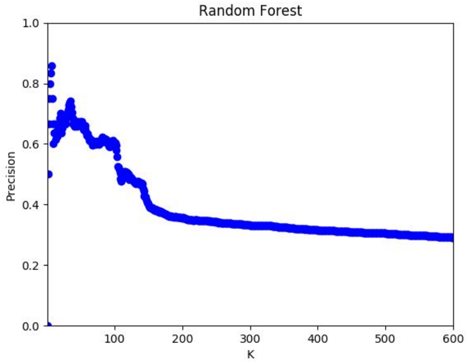
Label = 1

Standard Models

For our initial approach, we used a supervised learning technique with bag of words & TF-IDF (optional).
Below are the top 2 performing models.

Top Models	**acc**	**roc/auc**	**prec**	**recall**	**f1**
Logistic Regression	0.75	0.66	0.68	0.75	0.66
Random Forest	0.74	0.60	0.66	0.74	0.67





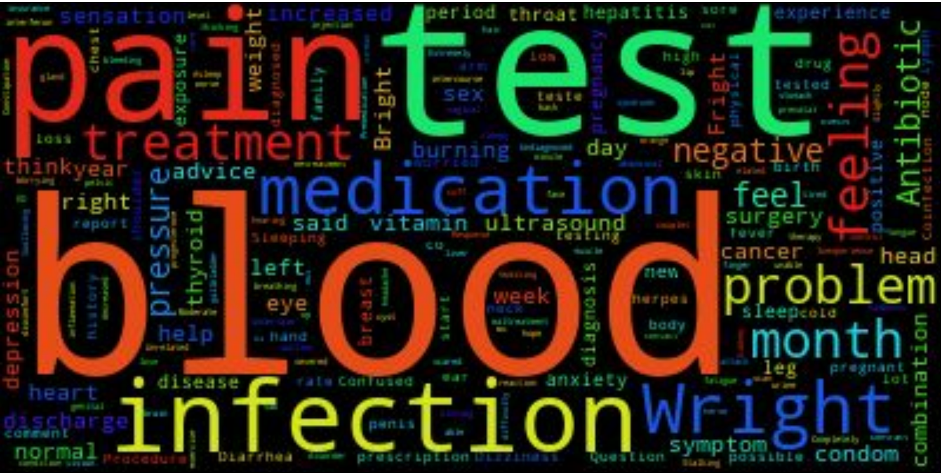
Standard Feature Analysis

- Beginning: **571,045 features**
- Feature Selection (chi-squared):
  - p-value < 0.05: 29,821 features
  - p-value < 0.01: 14,628 features
  - [no difference in model performance]

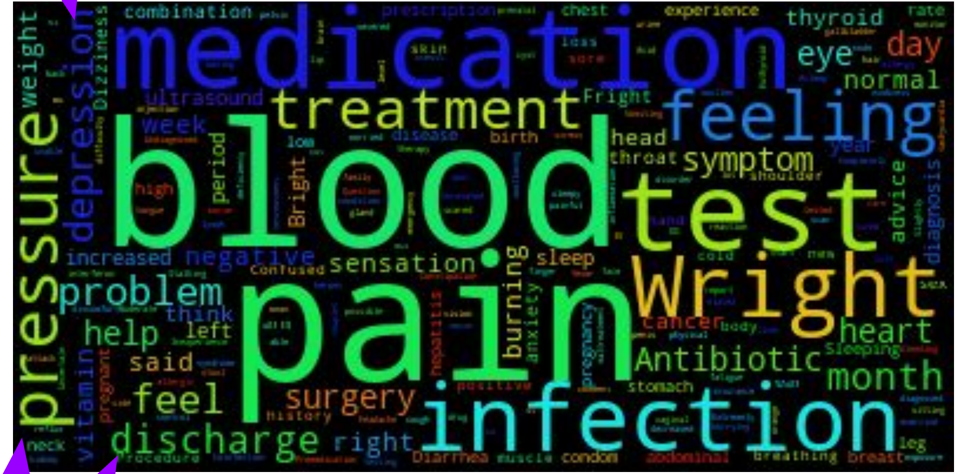
- Top 20 features
  - Back, **chest**, condom, doctor, feel, feeling, go, **heart**, **hiv**, like, normal, **pain**, risk, sex, side, sometimes, started, stomach, symptoms, **worse**

UMLS Word Frequency Analysis (no collocation)

Label = 0



Label = 1

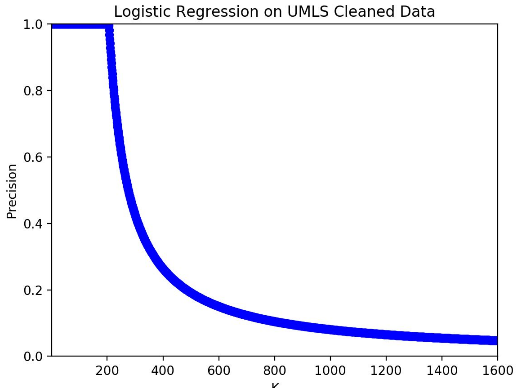


Incorporated QuickUMLS to identify medical concepts for each question

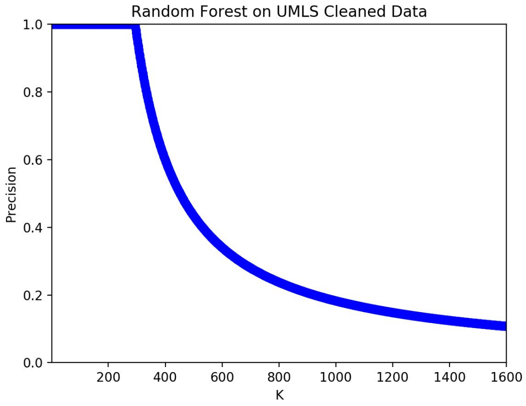
UMLS Models

We used a supervised learning technique with bag of words & TF-IDF. Below are the top 2 performing models.

Logistic regression has higher precision, but random forest has a better precision at k curve.



Top Models	**acc**	**roc/auc**	**prec**	**recall**	**f1**
Logistic Regression	0.75	0.66	**0.68**	0.75	0.66
Random Forest	0.74	0.60	0.66	0.74	0.67



UMLS Feature Analysis

- Beginning: 29,134 features
- Feature Selection (chi-squared, p-value < 0.05): 7,801 features
  - Decreasing p-value → no difference in model performance
    - p-value < 0.01: 5,616 features
    - p-value < 0.001: 4,201 features
- Top 20 features
  - abdominal, **breathing**, **chest**, condom, condoms, constipation, **depression**, discharge, dizziness, dizzy, **dog**, exposure, feeling, **heart**, **medication**, medications, **pain**, pressure, sex, stomach

Top 20 UMLS Features: Dog?

- Bivariate Relations
  - Dog run
  - Dog allergy
  - Dog bite
  - Dog companion
  - Dog day
- Dog community group
  - Questions for vets

Conclusions, Lessons & Future Work

Conclusions

1. UMLS models performs better than standards models
   a. Random Forest with UMLS performs the best
2. But, precision still has room for improvement

Lessons Learned

1. "Occam’s Razor"
   a. Sometimes the simplest model works the best
2. When working with big data, runtime is the bottleneck

Future Work

1. Continue training models to achieve higher support
2. Implement the neural network models (LSTM, CNN)

Thank you

References

1. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and E-mail for Health Care Information Results From a National Survey. *JAMA*. 2003;289(18):2400–2406. doi:10.1001/jama.289.18.2400
2. Lin, T. C., Lai, M. C., & Yang, S. W. (2016). Factors influencing physicians’ knowledge sharing on web medical forums. *Health Informatics Journal, 22*(3), 594-607. doi:10.1177/1460458215576229
3. Woo, J., Lee, M., & Ku, Y. (2012). The Informative Support and Emotional Support Classification Model for Medical Web Forums using Text Analysis. *Journal of the Korea Society of IT Services, 11*(Sup), 139-152. doi:10.9716/kits.2012.11.sup.139
4. Bader JL, Theofanos MF. Searching for cancer information on the Internet: analyzing natural language search queries. J Med Internet Res [serial on the Internet]. 2003 October–December [accessed February 2, 2007]; 5:[about 17 p.]. Available from: http://www.jmir.org
5. Cole, J., Watkins, C., & Kleine, D. (2016). Health Advice from Internet Discussion Forums: How Bad Is Dangerous? *Journal of Medical Internet Research, 18*(1). doi:10.2196/jmir.5051
6. Ahmad, F., Hudak, P. L., Bercovitz, K., Hollenberg, E., & Levinson, W. (2006). Are Physicians Ready for Patients With Internet-Based Health Information? *Journal of Medical Internet Research,8*(3). doi:10.2196/jmir.8.3.e22
7. Penson, D. F. (2016). Re: Changes in Burnout and Satisfaction with Work-Life Balance in Physicians and the General US Working Population between 2011 and 2014. *The Journal of Urology, 195*(5), 1568. doi:10.1016/j.juro.2016.02.051