

Analysis of a Pulmonary Health Screener

Christian Infante & Zachary Sather

Problem

- Pulmonary disease is a leading cause of death worldwide
- Much of the burden falls on developing countries (more than 90% of COPD deaths occur in low-income and middle-income countries)
- Need for high-quality and affordable solutions

The Tool

- Developed at MIT Mobile Technology Lab
- Mobile-based pulmonary health screener
- Questionnaire, peak flow meter, lung sound recordings
- Questionnaire performs the best
- Good classification of asthma and COPD

Motivations

- Motivation 1: subpopulation analysis
- Motivation 2: address allergic rhinitis (comorbidity)
- Motivation 3: shorten the questionnaire as much as possible

Data

- Responses to 70 questions
- 330 patients
 - 47 asthma
 - 35 COPD
 - 29 allergic rhinitis
 - 56 asthma + allergic rhinitis
 - 11 COPD + allergic rhinitis
 - 130 no pulmonary disease

K-means Clustering

	Cluster 1 (n=69)	Cluster 2 (n=122)	Cluster 3 (n=5)	Cluster 4 (n=112)
Age (yrs.)	62.49	41.11	44.2	50.71
Breathlessness	92.8	23.8	80.0	89.3
Cough (%)	100.00	0.00	100	83.93
Chest pain (%)	7.25	3.28	40.00	3.57
Fever (%)	0.00	0.00	100.00	0.00
Nasal symptoms (%)	0.00	2.46	80.00	97.32
Loss of appetite	14.5	4.1	80.0	5.4
Family history of COPD (%)	4.35	0.82	0.00	1.79
Family history of allergies (%)	26.09	23.77	20.00	41.96
Smoke (%)	37.68	12.30	20.00	27.68

K-means Clustering

- Cluster 1: older, smoker, chest pain (only 35% COPD)
- Cluster 2: less breathlessness, no cough (84% no pulmonary disease)
- Cluster 3: outliers, fever
- Cluster 4: nasal symptoms, family history of allergies (79% allergic rhinitis)

Factor Analysis

- Tried to explain variance in data from 15 latent factors
- Ended with 9 main latent factors (from 46 questions)
 - Nasal symptoms
 - Coughing
 - Breathlessness
 - Fever
 - Sputum abnormalities
 - Chest pain
 - Tobacco use
 - Biomass cooking
 - Throat symptoms

Logistic Regression (L1 Penalty)

- Multi-class (COPD, asthma, healthy)
- Using all questions, average score: 0.83
- Found top features for each class
 - COPD - Cigarettes/day, Smoking, Age
 - Asthma - Family history of allergies, Sticky sputum, Weight
 - Healthy - Age, Breathlessness, Fever sweating
- 12 questions achieve same average score

Logistic Regression (L1 Penalty)

- Allergic Rhinitis only
- Using all questions, average score: 0.97
- Top features
 - Nasal symptoms, Nasal discharge, Breathlessness
- 12 questions achieve same average score

Decision Tree Classification

- Found features with weights
 - Took the top 6 features (0.86)
 - Breathlessness
 - Onset (yrs.)
 - Age
 - Productive cough (phlegm)
 - Periods free of cough
 - Number of Cigarettes per day
 - Min. features: 12
- For Allergic Rhinitis
 - Top 6 Features (0.97)
 - Nasal symptoms & frequency
 - Nasal discharge
 - Nose: itchy, blocked
 - Breathing through mouth
 - Min. features: 12