Analyzing Influenza-like Illness with Unsupervised Machine Learning

Monique Bouvier PhD1, Mary Barger PhD1, David D’Ambrosio PhD, John Arnold MD2
1University of San Diego, San Diego CA; 2Naval Medical Center, San Diego CA

Introduction

- Acute respiratory infection outbreaks are a significant source of non-battle injury among otherwise healthy military personnel.
- Influenza-like illness (ILI) is the most common cause of healthcare visits with annual cost >$12 billion.
- Approximately 20% of samples from those with ILI were positive for influenza.
- Prior studies were unable to predict viral types from symptoms and few examined symptoms longitudinally.
- No studies have examined ILI symptoms through an unsupervised machine learning analysis.

Aims

- Identify if symptom presentation over the course of influenza-like illness (ILI) can predict virus type in a healthy military population using unsupervised machine learning.
- Identify sub-populations with similar symptom experience.

Unsupervised Machine Learning

- Process detects patterns in data by clustering samples into groups based on similarity. The technique K-means clustering is illustrated below.

Methods

Study Design – Data from Acute Respiratory Infection Consortium longitudinal study of ILI at 5 US military facilities, 2009-14.
- Healthy military members, retirees, and dependents with ILI symptoms presenting within the last 72 hours.
- ILI definition: Fever (>100.4°F) with at least one respiratory symptom.
- Prospective symptom data collected at visits 0,3, and 7 using a symptom list graded on a standardized 4-point scale.
- Categories of symptom severity: Upper respiratory, Lower respiratory, Systemic, Gastrointestinal
- Nasal swabs for typing virus.

Analysis - Unsupervised machine learning algorithm k-means clustering.
1: All subjects were clustered according to individual symptom severity scores (days 0, 3, 7).
2: Subjects separated into groups of virus type and symptom expression were clustered. Subjects attributes (age, sex, ethnicity, smoking history, BMI, and Military status) within clusters compared.

Conclusion

- First study to analyze individual ILI symptom scores through unsupervised machine learning technique.
- A person’s characteristics may influence how they experience physical symptoms and severity of an illness, limiting usefulness of physical severity scores.
- Healthcare providers need to take into account outside factors like environmental, biological, or social that may influence symptom severity.
- Data from the ‘no virus’ group was difficult to use for comparison since the subject most likely had a bacterial or viral illness not detected from biologic nasal swab.
- Secondary analysis of data limited analysis; unable to assess all constructs of SMT model below. Symptom severity instrument only captured physical symptoms.

Future Research

- Perform further analysis with other clustering methods and compare results to traditional analysis techniques.
- Study symptom experience of ILI from the perspective of the Symptom Management Theory (SMT)

Attributes by Symptom Clusters

Future Research

- Perform further analysis with other clustering methods and compare results to traditional analysis techniques.
- Study symptom experience of ILI from the perspective of the Symptom Management Theory (SMT)

Acknowledgements and Disclaimer

The Jonas Center for Nursing and Veteran Healthcare.
Partial support for this work was provided by the Department of Defense Global Emerging Infections Surveillance (GEIS) program. Additional support has been provided in whole, or in part, with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, under Inter Agency Agreement Y1-AI-50727 to Uniformed Services University of the Health Sciences. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University, Department of the Navy, Army, Department of Defense, the U.S. Government, nor the Henry M. Jackson Foundation.