100 years of sepsis: Using topic modeling to understand the historical themes surrounding sepsis

Doran Bostwick1,2, Kelly Peterson3,4, Barbara Jones1,2, Matt Samore2,4, Robert Paine1,2, Makoto Jones2,4

1 Pulmonary and Critical Care, University of Utah, 2 VA Salt Lake City Healthcare System, 3 Division of Epidemiology, University of Utah, 4 Infectious Disease, University of Utah

BACKGROUND / OBJECTIVES

- Medical research publications have increased at an exponential rate. However, our capacity to absorb and understand large amounts of data is limited.
- Synthesizing prior work to inform future research is critical for the continued advancement of medical knowledge.
- Topic modeling is a statistical model that generates topics based on relatedness of words.

Objective:

Infer the discourse surrounding sepsis up to the present by distilling large amounts of information into its elemental themes using topic modeling.

METHODS

- Extracted all abstracts from Pubmed containing the terms “sepsis”, “septic shock”, and “septicemia” from the years 1890-2018.
- Using topic modeling approaches based on Latent Dirichlet Allocation we:
  1) Trained dynamic models to identify five topics from the corpus
  2) Developed a static topic model for the last 5 years
- Conducted a thematic analysis of topics across publication periods by examining the 30 most frequent words in each topic for each decade and also for the 5 topics generated from the last 5 years.
- Compared the frequency of each topic over the first and second halves of the century.
- Compared the respective themes and their relatedness
- Over the last 100 years, the focus shifted from the pathogen to the host response both from a cellular and physiologic perspective. Additionally, the diversity of topics in the literature has increased.

In the last 5 years, biomarkers, early recognition and system management emerged as new themes.

Reasons for this may include:

i. Evolution of scientific tools, treatments and statistical abilities
ii. Increased focus on the cost of healthcare
iii. Incorporation of the individual host response into the disease model

Topic modeling is a tool that can enhance understanding of a large complex field and provide a framework to inform future research.

ACKNOWLEDGEMENTS

- Project funded by Department of Veterans Affairs, HSR&D
- Dr. Bostwick is supported by the 372 Grants from the NIH
- This project has used data resources from VA Informatics and Computing Infrastructure (VINCI) at the University of Utah, Salt Lake City, UT
- The views expressed are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

CONTACT: Doran Bostwick, University of Utah, Salt Lake City, UT

Presented at the 2018 ID Week Annual Conference