Leveraging Language Models Generated Embeddings for Topic Modeling and Survival Analysis of Obstructive Sleep Apnea Patients

Himanshu Yadav[1], Destinee Morrow[2]

[1]University of Florida, [2]Lawrence Berkeley National Laboratory

ABSTRACT

This study leverages the MIMIC-IV database to analyze clinical narratives from discharge notes of patients diagnosed with Obstructive Sleep Apnea (OSA) using BERTopic with GatorTron embeddings for topic modeling and survival analysis. We identified key themes such as "Symptoms and Diagnosis," "Treatment and Medication," and "Follow-up and Monitoring." Temporal analysis revealed the evolution of these themes over time. Kaplan-Meier is used to examine the impact of various covariates on patient outcomes, providing insights into the clinical management of OSA and significant factors affecting patient survival.

BACKGROUND INFO

Obstructive Sleep Apnea (OSA) is a prevalent sleep disorder that affects breathing during sleep and can lead to severe health issues like cardiovascular disease and hypertension. Managing OSA is essential for improving patient outcomes and reducing healthcare burdens.

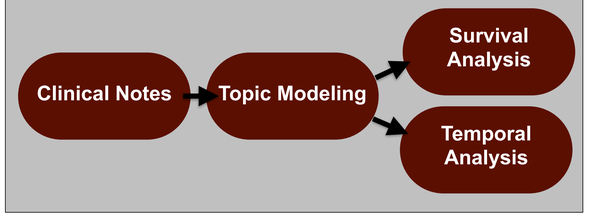
MIMIC-IV Database: The Medical Information Mart for Intensive Care IV (MIMIC-IV) is a vast, de-identified database containing patient data from critical care units. It includes diverse data types such as demographics, vital signs, lab results, medications, and clinical notes, offering extensive resources for health condition analysis.

BERTopic: BERTopic is an advanced topic modeling technique that uses transformer-based language models to extract themes from text by clustering embeddings.

GatorTron: GatorTron is a specialized language model fine-tuned on clinical texts, enabling it to generate high-quality embeddings that capture medical terminology and context effectively.

Research Hypothesis

Advanced topic modeling using Language Models, combined with survival analysis, can uncover critical clinical themes and significant factors that affect patient outcomes in OSA management!



OSA Patient Demographics

- **Total Number of Patients with OSA:** 14,915
- **Male Patients:** 8,918
- **Female Patients:** 5,997

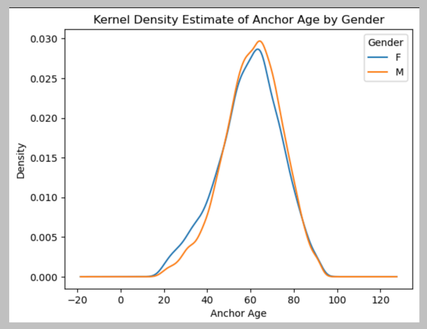


Figure 1: *Kernel Density Estimation of Age for Male and Female Patients diagnosed with OSA.*

Topic Modeling

Initial Analysis: Processed 51,565 clinical notes, identifying 350 topics (Figure 2).

Refinement: Applied hierarchical topic matching (thresholds 0.8 and 0.85) to merge overlapping topics, resulting in 166 distinct categories.

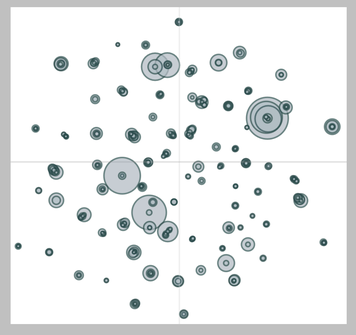


Figure 2: *This plot visualizes the initial 350 topics identified from the clinical notes, with overlapping regions indicating common themes. The size of each circle corresponds to the number of notes within that topic.*

Prominent Topics: Analyzed and identified the top 8 topics based on their frequency in the notes (as shown in Figure 3).

Visualization: Bar plots in Figure 3 display key keywords for each prominent topic, highlighting their relevance within the clinical notes and major themes in OSA discharge summaries.

Key Themes in OSA Notes (Figure 3):
- Topic 0: Neurological symptoms
- Topic 1: Cardiac issues
- Topic 3: Surgical procedures
- Topic 4: Puncture procedures and dietary considerations
- Topic 5: Psychiatric concerns
- Topic 6: Dehydration symptoms

Themes in Clinical Notes

Figure 3: *Visualization of the top 8 topics after performing topic modeling and merging similar topics. Each bar plot represents a topic, with the most significant words and their scores depicted to signify their importance within the topic.*

Temporal Analysis

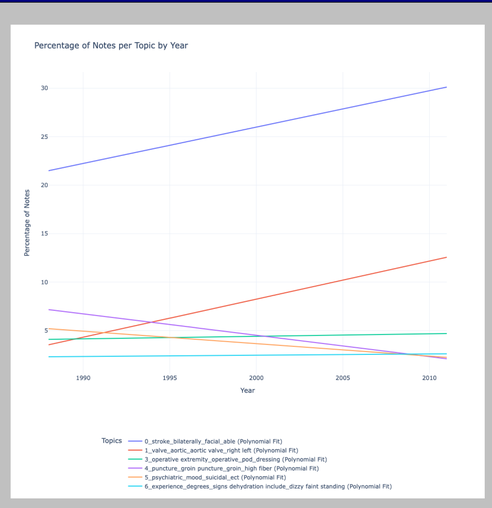


Figure 4: *Temporal distribution of clinical notes across identified topics from 1987 to 2012. 2000 is the year when the patient was first diagnosed with OSA. Each line represents the percentage of notes attributed to a specific topic per year*

Objective: Examine how clinical themes evolved over time using the diagnosis date of OSA patients as a reference point.

Methodology: Standardized diagnosis dates to January 1, 2000, 00:00:00 and analyzed the distribution of topics in clinical notes by year, as shown in Figure 4.

Findings:
- **Increase in Cardiac Issues:** Topic 1 shows a significant rise over time, emphasizing growing cardiovascular concerns.
- **Neurological Symptoms:** Topic 0 steadily increases, indicating consistent focus on neurological issues.

Implications: The evolving focus highlights an increased understanding of the long-term health impacts of OSA and the need for targeted management of cardiovascular complications.

Survival Analysis

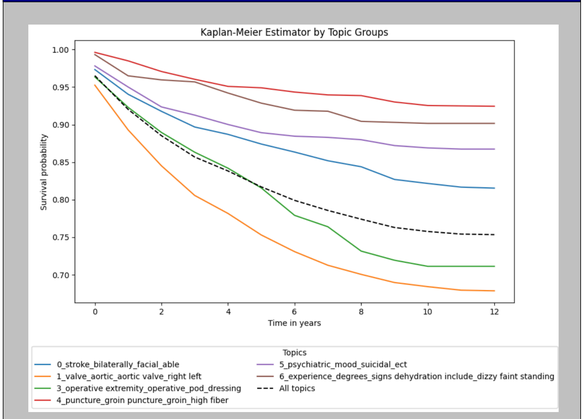


Figure 5: *Kaplan-Meier survival curves for six topic groups derived from OSA patient discharge notes, showing varying survival probabilities over 12 years based on different clinical themes.*

Main takeaways from Kaplan-Meier survival curves (Figure 5):
- **Highest Survival:** Topic 4 (Puncture/Dietary Considerations) remains consistently high.
- **Increasing Trends:** Topics 6 (Dehydration Symptoms) and 5 (Psychiatric Issues) show positive trends in survival.
- **Moderate Survival:** Topic 0 (Neurological Symptoms) display steady declines.
- **Lowest Survival:** Topics 3 (Surgical Procedures) and 1 (Cardiac Issues) indicates the highest mortality risk.

These patterns highlight the differing impacts of clinical themes on survival rates, guiding tailored OSA patient management.

Conclusion and Future Work

Conclusion: Our study effectively identified crucial clinical themes and factors that influence patient outcomes in the management of Obstructive Sleep Apnea (OSA). By analyzing a vast array of clinical notes, we have highlighted significant patterns that can inform better clinical practices and treatment approaches.

Future Work: Moving forward, we aim to refine our analytical models further, incorporate more comprehensive datasets, and develop tailored treatment strategies. These efforts will be directed towards enhancing patient care and clinical outcomes for OSA patients, ultimately aiming to reduce the healthcare burden associated with this prevalent condition.

ACKNOWLEDGMENTS

This poster template was created by Workforce Development & Education at Berkeley Lab.