Product design / development / engineering | Healthcare / Pharmaceuticals | Product design and testing | Unsupervised Learning

Analysis of Respiratory Time Series Data for Breathing Discomfort Detection

BUSINESS PROBLEM

Patients with obstructive sleep apnea (OSA) often cite discomfort as a reason for discontinuing continuous positive airway pressure (CPAP) therapy. Many healthcare providers do not have time or awareness to proactively intervene with adjustments that would improve therapy adherence. ResMed aims to understand which features in high-resolution respiratory data can be used to detect critical respiratory events such as breathing discomfort. The company also seeks to understand more effective ways to convey data insights to healthcare providers for better decision making.

APPROACH

Time-series clustering was applied to identify groups of patients who show similar patterns in their breathing characteristics. Clusters were inspected to understand the relationship between breathing signals and patient reports of discomfort on low pressures, and other variables of interest. Statistical tests were used to evaluated whether differences in variables across groups were significant.

DATA SOURCES

Several datasets are available and include the following: patient demographic information, device usage data, and high-resolution time series respiratory data. Data has previously been cleaned and de-identified and are in tables in ResMed's central database. Other data sources include patient reported data on air quality and quantity during the sleep onset period of therapy.

Data Types and Format

Patient data is stored in ResMed's database and tables are accessible via query. Qualitative feedback from patients was gathered via a survey and consolidated into summary reports.

Author: Shelby Unger

IMPACT

With the rise of cloud-connected CPAP devices, the sleep medicine industry is shifting towards products that offer a higher level of automation and personalization. ResMed is working to understand how it can leverage different sources of data to provide new products for health care providers and patients that are aligned with this trend. This project evaluates how effective high resolution respiratory data is for identifying patients who may be experiencing breathing discomfort or stuffiness at low pressure levels during the sleep onset period of therapy. The time series clustering approach enables the automatic segmentation of patients who have similar breathing patterns. By conducting statistical testing on variables relating to demographic characteristics, signal features, usage, and reporting of breathing discomfort, ResMed is able to assess which features are independent of the clustering and which are not. As a whole, this allows ResMed to understand which features in their data may be useful for predicting which patients are experiencing breathing discomfort and understand where they may need to invest in the collection of additional data to improve model accuracy.

DRIVERS

The advent of the cloud-connected CPAP machine has enabled ResMed to gather large amounts of data relating to therapy usage and health outcomes across different demographic groups. Today, the organization is investing heavily into new ways that it can leverage this information to improve patient outcomes. ResMed as a whole is focused on data-driven solutions for remote patient management, personalized treatment, and virtualized care.

BARRIERS

Industry regulations make it difficult to gather new types of data quickly. Self-reported data from patients also introduces challenges such as recall bias, which impacts model performance. Biological variability also means that respiratory events may present differently in each patient, making the model development cycle longer and more complex.

ENABLERS

ResMed has a data-driven strategy with a dedicated data science and artificial intelligence (DSAIL) team. The organization embraces data-driven solutions. It supports the data collection and experimentation necessary to generate data insights and productize machine learning.

ACTIONS

I collaborated closely with ResMed's DSAIL team and domain experts to understand the data the was available and its meaning in the context of the use case being explored. I then tested several modeling approaches, selected the one that yielded the best segmentation of patients, and assess the significance of differences in variables across the groups with statistical testing.

INNOVATION

This project applies a time-series clustering approach to categorize respiratory data gathered throughout the night by ResMed's AirSense 11 machines with the goal of understanding how different patterns may relate to patient reports of breathing discomfort at low pressure levels.

IMPROVEMENT

This analysis enabled ResMed to understand what features may be useful for identifying breathing discomfort in patients and understand limitations in their current data with regards to this application. Several recommendations were made for how ResMed might consider collecting additional data to enable accurate detection of breathing discomfort in the future.

BEST PRACTICES

Test several types of models and model parameters to confirm which suits your objective and data best. Work closely with subject matter experts throughout the process to confirm and understand results and insights.

OTHER APPLICATIONS

Clustering has been shown to be a useful tool for discovering patterns in labeled and unlabeled data sets for applications within the medical field and beyond. Time series clustering could be applied to other biological signals to understand different subgroups within the population, or to any other type of application with sequential data.