Activity Recipes for High Sleep Efficiency using Wearable Device Data

Saksham Goel | goelx029 | 5138568
Mentor: Meghna Singh, Professor: Jaideep Srivastava
Department of Computer Science, University of Minnesota Twin Cities

Abstract

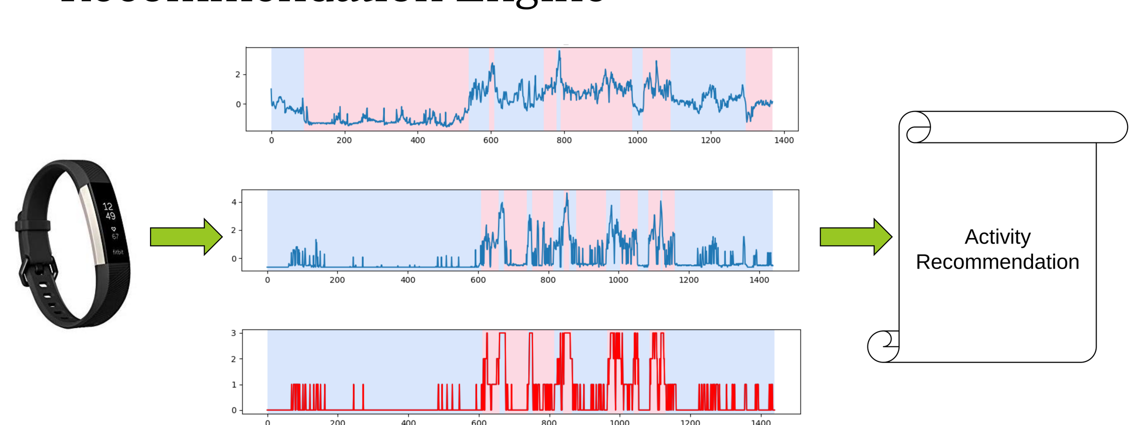
My project tries to address whether day to day activities affect quality of sleep, and if there are activity recipes that can be learned for making recommendations to users to ensure good sleep. This is achieved by applying clustering techniques over intraday time series data collected from wearable devices to extract activity recipes for high sleep efficiency, which will then be supplied to an activity recommendation engine. Various distance metrics are evaluated for their effectiveness in processing health time series data.

Introduction

Significance: Inadequate sleep negatively affects mental and physical well-being and exacerbates health problems such as diabetes, depression, cancer and obesity.

Project Objective:

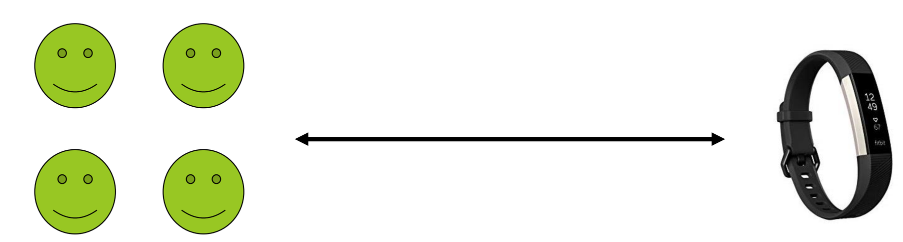
- Extract actionable knowledge of activity recipes for high sleep efficiency using low cost, ubiquitous, low fidelity data for health and wellness from wearable devices like Fitbit.
- Supply activity recipes learned to an Activity Recommendation Engine



Tasks:

- Heart rate data to cluster biologically similar people
- K-Means Clustering to find similarity among good sleep instances and learn best activity recipes for high sleep efficiency.

Methodology



Fitbit Web API

Data Collection Script

FITBIT DB

Intraday Data (Heart, Sleep, Activity, Steps)

LAB DB

Data Validation and Cleaning

Feature Extraction

K-Means Clustering

Highest Purity = Lowest overall Gini index.

$$Gini = 1 - \sum_{i=1}^{n} P(x_i)^2$$

Optimal number of clusters

Sub Clusters Clustering

- K-Means on each of the clusters individually.
- Features: Activity level summaries of the day.
- Similar distance metric

Good Sleep Activity Recipes Extraction

$$Sleep\ Ratio = \frac{Total\ Minutes\ Asleep}{Total\ Minutes\ in\ Bed} = 1 - \frac{Minutes\ Awake\ in\ Bed}{Total\ Minutes\ in\ Bed}$$

$$Cluster\ Ratio = \frac{Number\ of\ Good\ Sleep\ Records}{Number\ of\ Bad\ Sleep\ Records}$$

- Good sleep = Sleep Ratio $\geq$ 0.90
- Good Sub Clusters = Cluster Ratio $\geq$ 2

<u>Data Collection:</u> 4 people over a time period of 4 months.

<u>Data Cleaning:</u> Removing NaN values using means found in hourly segments corresponding to each user.

<u>Feature Extraction:</u> Smoothing using a seasonal decomposition of moving window length 10 minutes

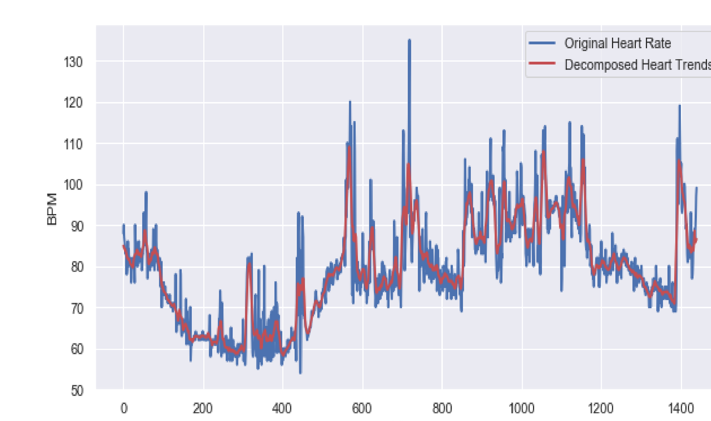


Figure: Trends extracted from the raw heart rate data

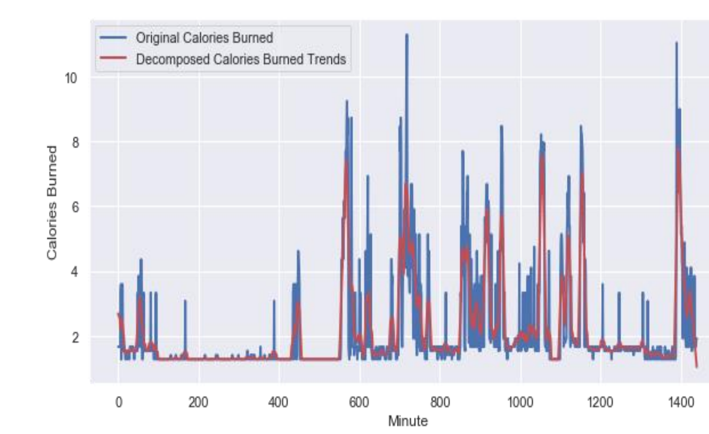


Figure: Trends extracted from the raw calories rate data

Distance Metrics:

- L-1 Norm = $\sum_i |x_i - y_i|$
- L-2 Norm = $\sqrt{\sum_i |x_i - y_i|^2}$
- Dynamic Time Wrapping
- Correlation = $\frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{\sum X^2 - n\bar{X}^2}\sqrt{\sum Y^2 - n\bar{Y}^2}}$
- K-L Divergence = $\sum_i x_i \times \log(\frac{x_i}{y_i})$

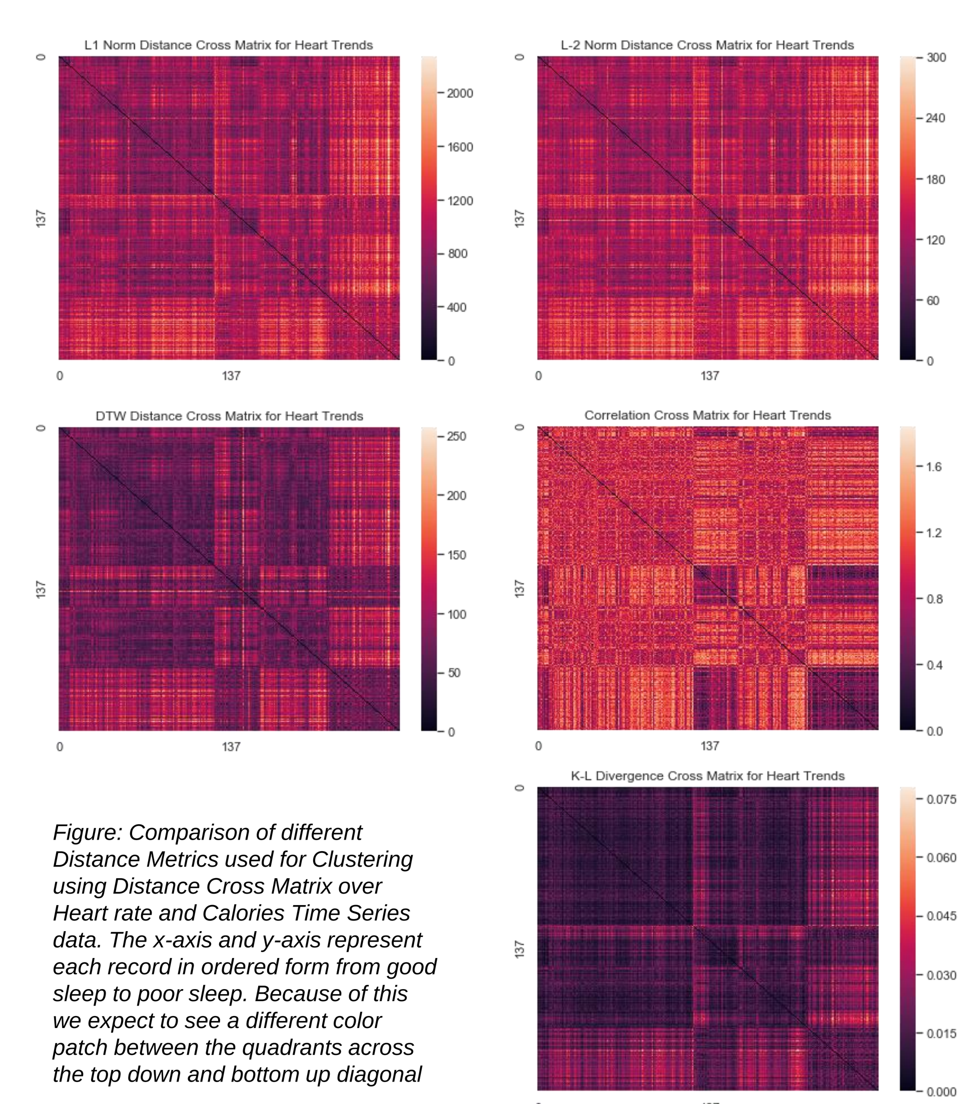


Figure: Comparison of different Distance Metrics used for Clustering using Distance Cross Matrix over Heart rate and Calories Time Series data. The x-axis and y-axis represent each record in ordered form from good sleep to poor sleep. Because of this we expect to see a different color patch between the quadrants across the top down and bottom up diagonal

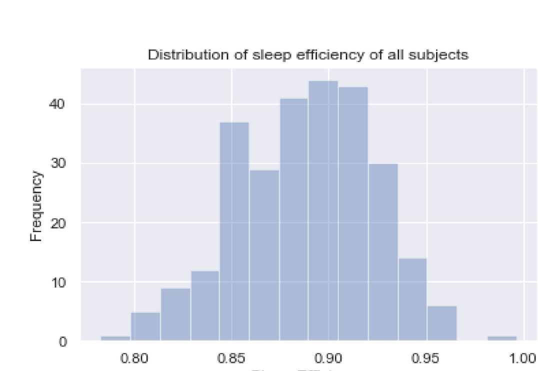


Figure: Distribution of the sleep efficiency scores in the dataset

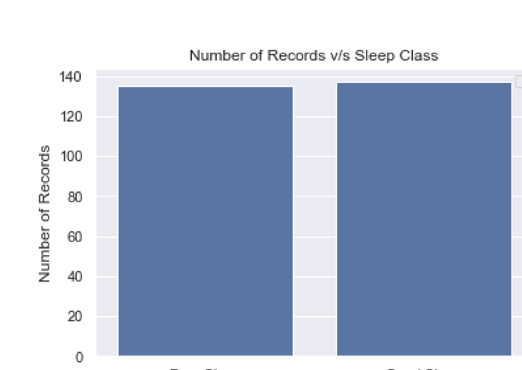


Figure: Number of records for Good v/s Poor sleep

Results

Distance Metric	Purity of Clustering	Avg. Purity of Activity Recipe Clusters
L-1 Norm	0.5473	2.13
L-2 Norm	0.5848	2.34
DTW	0.6372	2.59
Correlation	0.6516	3.12
K-L Divergence	0.7006	4.18

Table: Comparison of different distance metric functions over clustering purity

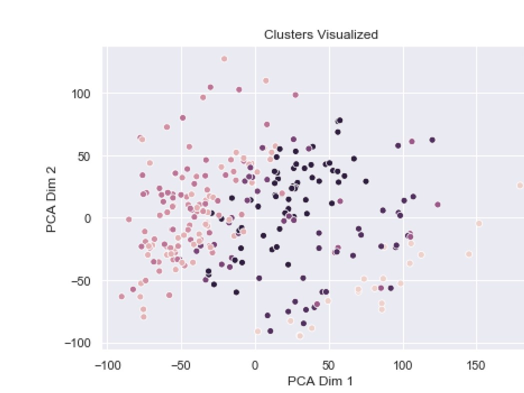


Figure: Cluster Assignments of heart rate trends using K-L Divergence

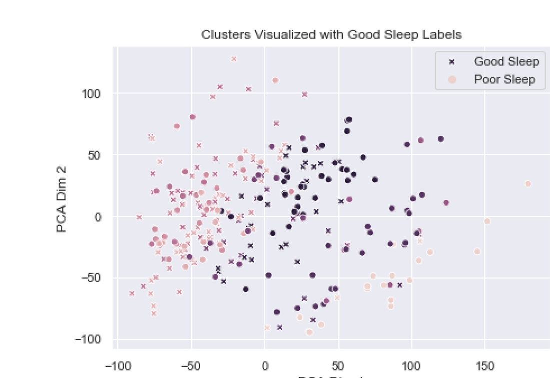


Figure: Cluster Purity of Sleep Labels using K-L Divergence over heart rate trends

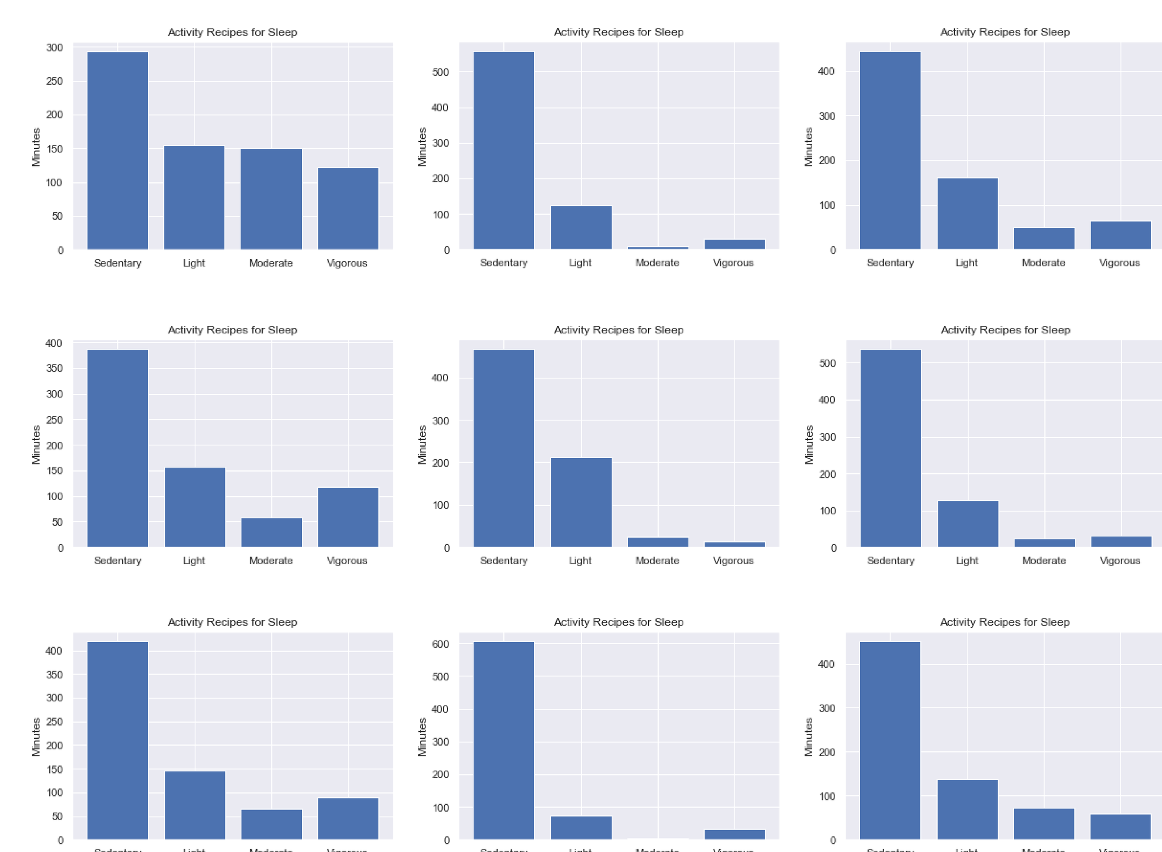


Figure: Activity Recipes for good sleep found using K-L Divergence distance metric

Conclusions

- Good sleep nights have similar preceding daily activities, while poor sleep can vary
- Clustering over heart rate lead to clusters containing varied activity level helping learn myriad different activities from same cluster
- Among the distance metrics, K-L divergence performs best in finding the purest cluster w.r.t sleep efficiency and produces meaningful and varied activity recipes