Multimodal Ambulatory Sleep Detection Using Recurrent Neural Networks


1Media Lab, Massachusetts Institute of Technology, Cambridge, MA,
2Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA,
3Sleep Health Institute and Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA,
4Division of Sleep Medicine, Harvard Medical School, Boston, MA.
Conflict of Interest Disclosures for Speakers

1. I do not have any relationships with any entities producing, marketing, re-selling, or distributing health care goods or services consumed by, or used on, patients, **OR**

2. I have the following relationships with entities producing, marketing, re-selling, or distributing health care goods or services consumed by, or used on, patients.

<table>
<thead>
<tr>
<th>Type of Potential Conflict</th>
<th>Details of Potential Conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grant/Research Support</td>
<td></td>
</tr>
<tr>
<td>Consultant</td>
<td></td>
</tr>
<tr>
<td>Speakers’ Bureaus</td>
<td></td>
</tr>
<tr>
<td>Financial support</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

3. The material presented in this lecture has no relationship with any of these potential conflicts, **OR**

4. This talk presents material that is related to one or more of these potential conflicts, and the following objective references are provided as support for this lecture:

1. 
2. 
3.
Motivation

Polysomnography (PSG)
Impractical for long-term home use

Actigraphy + Sleep Diary
Requires significant effort of users to maintain accurate diaries, and of researchers to check the diary entries for anomalies

There is a need for tools to enable accurate long-term evaluation of sleep timing and duration in daily life with less burden on users and researchers.
Data

- 5580 days of **multimodal data** from a wrist sensor and an Android phone
- 186 undergraduate students, 30 days each

Wrist Sensor
- Skin conductance (SC)
- Acceleration (ACC)
- Skin temperature (ST)

Phone
- Call
- SMS
- Location
- Screen

Time

Labels of sleep/wake:
- Human scored actigraphy with sleep diaries based on a previously established method (Barger et al., 2014)
  - Resolution: 1 min -> 1 day = 1440 labels
### Features

SC is more likely to have periods of high frequency activity called “storms” during NREM2 and SWS sleep.

Movement index = \( \frac{(\text{var(latitude)} + \text{var(longitude)})}{2} \)

<table>
<thead>
<tr>
<th>Source</th>
<th>Modality</th>
<th>Feature variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrist sensor</td>
<td>Skin conductance (SC)</td>
<td>Mean, SD, power within 0.1, 0.1-0.2, 0.2-0.3, 0.3-0.4, and 0.4-0.5 Hz bands, the number of SC responses, storm flag, elapsed time since a storm started</td>
</tr>
<tr>
<td></td>
<td>Acceleration (ACC)</td>
<td>Mean, SD</td>
</tr>
<tr>
<td></td>
<td>Skin temperature (ST)</td>
<td>Mean, SD</td>
</tr>
<tr>
<td>Phone</td>
<td>Screen</td>
<td>Screen was on, the time the screen was turned on</td>
</tr>
<tr>
<td></td>
<td>SMS</td>
<td>Sent a message</td>
</tr>
<tr>
<td></td>
<td>Call</td>
<td>On a call, missed a call</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>Movement index, connected to WiFi, connected to cellular nets</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>Elapsed minutes since 12:00 AM</td>
</tr>
</tbody>
</table>
Methods

Sleep detection:
Bidirectional long short-term memory neural network model

\[ y_t = \text{Sleep Probability} \]

Sigmoid

Fully-connected

0.2 Dropout

Concatenation

LSTM 32 \[ \cdots \] LSTM 32 \[ \cdots \] LSTM 32 \[ \cdots \] LSTM 32 \[ \cdots \] LSTM 32

\[ x_{t-29} \] \[ \cdots \] \[ x_{t-1} \] \[ x_t \] \[ x_{t+1} \] \[ \cdots \] \[ x_{t+29} \]
Methods

Sleep episode onset/offset detection:
Bidirectional long short-term memory neural network model
+ Peak detection

$y'_i = \text{Sleep Offset Probability}$

![Diagram showing the neural network model for sleep episode detection.](image)

Offset probability

Ground truth

Time / min

0 2000 4000 6000 8000 10000

0 0.02

SLEEPS 2017

BOSTON ★ JUNE 3-7
Results

Sleep/wake classification accuracy: **96.5%**
*(Acceleration + Skin temperature + Time)*

For each participant,
80% of days - training set,
20% of days - test set

| Sleep episode onset detection | F₁ scores: **0.86**, mean errors: **5.0 min** |
| Sleep episode offset detection | F₁ scores: **0.84**, mean errors: **5.5 min** |

<table>
<thead>
<tr>
<th>Feature combinations</th>
<th>Sleep detection accuracy</th>
<th>Sleep episode on/offset detection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without time</td>
<td>With time</td>
</tr>
<tr>
<td>Wrist sensor</td>
<td>95.9</td>
<td>96.5</td>
</tr>
<tr>
<td>Phone</td>
<td>76.7</td>
<td>89.1</td>
</tr>
<tr>
<td>Wrist + Phone</td>
<td>96.0</td>
<td>96.3</td>
</tr>
<tr>
<td>ACC + EDA</td>
<td>95.6</td>
<td>96.3</td>
</tr>
<tr>
<td>ACC + ST</td>
<td><strong>96.2</strong></td>
<td><strong>96.5</strong></td>
</tr>
<tr>
<td>EDA + ST</td>
<td>92.5</td>
<td>94.5</td>
</tr>
<tr>
<td>ACC</td>
<td>95.5</td>
<td>96.3</td>
</tr>
<tr>
<td>EDA</td>
<td>90.8</td>
<td>93.7</td>
</tr>
<tr>
<td>ST</td>
<td>86.7</td>
<td>90.7</td>
</tr>
</tbody>
</table>
Generalized to different participants

80% of participants - training set
20% of participants - test set

Participant 1

Participant 2

![Graph showing sleep detection accuracy for different participant data sets](image-url)
Real-time implementation
Conclusion

We showed

Sleep/wake classification accuracy: 96.5% with features from Acceleration + Skin temperature + Time

Sleep episode onset detection ($F_1$ scores: 0.86, mean errors: 5.0 min)

Sleep episode offset detection ($F_1$ scores: 0.84, mean errors: 5.5 min)

Our results indicate that long-term ambulatory sleep/wake records from large populations can be measured unobtrusively and accurately by exploiting the ubiquity of smartphones and wearable sensors and the power of deep learning.