SLEEP MOVEMENT DETECTION USING *K*-NEAREST NEIGHBORS AND DYNAMIC TIME WARPING

[1]VISHAL V, [2]C. UDAYA KUMAR

[1,2]Manipal Institute of Technology, Manipal, Karnataka, India
E-mail: [1]vishalv7197@gmail.com, [2]udayakumar97@gmail.com

Abstract- Analysis of sleep quality and sleep monitoring can serve as a diagnostic feature to check various sleep and psychiatric disorders. The quality of life of a large number of people is affected by diseases like sleep apnea, insomnia, narcolepsy etc. We propose a sleep movement detection algorithm that works accurately to classify sleep movement and non-movement data obtained from unobstructed pressure sensitive textile sheet called *Dozee*[1]that uses the Ballistocardiogram (BCG) technique to record signals. The data was collected from 8 subjects, each subject contributing 6 hours of sleep data. We utilized the K-nearest neighbors classification algorithm which is a non-parametric, classical supervised Machine learning algorithm, along with Dynamic time warping method to classify the time-series data and deduce inference from it. The proposed sleep movement detection algorithm achieved an overall accuracy of 98.17%.

Index terms- Machine learning, Time series classification, Sleep movement detection, Dynamic time warping

I. INTRODUCTION

Sleep plays a vital role in one's life. A person spends almost one-third of his lifetime in the act of sleeping. Sleep monitoring can aid in the diagnosis of a variety of sleep and psychiatric disorders. Sleep apnea, insomnia, narcolepsy, are all diseases that affect the quality of life of a large amount of population [1]. The quality of sleep we get during the night can affect our mental state and physical performance during the day. Sleep consisting of greater movement can be an indication of disrupted sleep and inferences such restlessness, psychological disorders, depression and many more can be derived. The quality of sleep is greatly enhanced in case of lower motor activity during the night. Our aim is to classify non-movement sleep patterns from sleep movement patterns in order to learn more about the correlation between quality of sleep and sleep motion and derive inferences from it [2]. There are several Supervised Machine Learning algorithms that can classify the Time series sleep data [3].

Polysomnography (PSG), also known as traditional sleep monitoring, is done with a wide array of sensors that is applied to the body of the patient. It can disturb the sleep of the patient. PSG is an intrusive method and is often done in sleep laboratories in the presence of trained technicians, hence it is usually expensive and not suitable for long-term sleep monitoring [4]. The increasing interest for automatic sleep monitoring is due to the fact that, in contrast to the PSG studies, these signals are measured with unobtrusive instruments. We will use a set of data recorded by a high-resolution pressure-sensitive textile sheet, called *Dozee* that uses the Ballistocardiogram (BCG) technique to record signals. This system compared with other techniques has many advantages: is unobtrusive, fits into users familiar sleep environment, is comfortable and is economical [5].

We performed Time-series Classification by using K-nearest neighbors and Dynamic Time Warping. The k-NN and DTW classification algorithm is one of the best and accurate ways to classify the Time Series data. The k-NN algorithm is a non-parametric method and a type of instance learning technique used for classification and regression [6]. It takes an unlabeled query observation and compares it to a population of labeled observations [7]. The query observation is classified by assigning the label which is most frequent among the k labeled samples (k is a user-defined constant) nearest to that query point. Dynamic Time warping is one of the algorithms in Time series analysis for measuring the similarity between two temporal sequences which may vary in speed [8]. This technique was heavily used for speech recognition in the 1980s. The DTW algorithm finds the optimum alignment between any two sequences of observations. It finds the alignment by warping the time dimension with certain constraints. A combination of k-NN and DTW serves as a good classification algorithm for Time series data. We will describe our Data collection and Extraction methodologies in section 2 and in section 3; we will provide an elaborate explanation of the classification method using k-NN and DTW. We would evaluate our experimental results from the classifier in section 4 and in section 5, we conclude by discussing the scope, limitations, drawbacks and future work pertaining to the algorithm.

II. DATA COLLECTION AND EXTRACTION

We acquired our data set consisting of signal values at a sampling rate of 250 Hz. The duration of a sleep data from each subject is 6 hours long. This data is further broken down into files of 4 minutes long. The data contains sensor values like humidity,

temperature, light conditions, time stamp and piezoelectric sensor values (pressure values). The piezoelectric sensor values (pressure values) are of interest to us to detect motion patterns during sleep [9]. It would be an interesting study to analyze the correlation between the condition values and the sleep motion; however, we would leave it for future work. Currently, we have access to data of 8 subjects with age ranging between 26 and 40 years. The raw signals obtained from the device consist of the heart, respiratory and motion artifacts. As sleep pattern consists of several movement patterns, we detect movement patterns at each second using supervised learning technique. The raw signal would be divided into epochs of 30 seconds and movement analysis is done in batches of 30 seconds.A binary classification is to be performed for each second of sleep data (‘0’ for non-movement and '1' for movement). The dataset is manually labeled for any movements in sleep data by verifying with the video capture of the subject's sleep. Hence, we create a labeled data set with labels '0' and '1' for non-movement and movement respectively for each second. The data set is split into training and test data sets. We allocate 70% of the data set for training the Classifier and the rest 30% for testing and validating the performance of the Classifier. For the performance evaluation, the labels are going to serve as ground truth for testing and validating our results. We would incorporate confusion matrix as our performance metric.

The Data obtained with help of *Dozee* is ASCII encoded, so the Data Extraction step is performed to extract sensor values and time stamps from the raw data. Base64 is used to decode the raw data and aggregate the values from each of the 4-minute files into a large CSV file so that it can be easier for us to process the data for further steps.

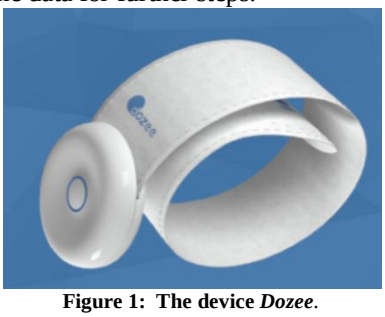

Figure 1: The device *Dozee*.

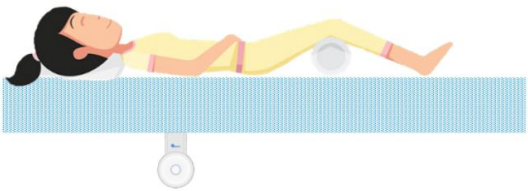

Figure 2: Data acquisition while Dozee under the mattress

III. THE k-NN AND DTW CLASSIFIER

The k-nearest neighbors algorithm is one of the classical Supervised Machine learning algorithms. It is an instance learning technique used for tasks such as classification and regression. The labeled data (training set) pertaining to a dataset is placed in Euclidean space. The task of the classifier is to classify an unlabeled query observation; hence it places the query observation in the Euclidean space in the pool of labeled data. The user defines the 'k' value which denotes the number the observations surrounding the query observation. A label is assigned to the query observation on the majority basis i.e. the label that has the maximum majority in the 'k' labeled observations pool (circle).

The Dynamic Time Warping is one the algorithms used in time series analysis to measure the similarity between two temporal sequences (here, time series data) which may vary in speed. It figures out an optimum alignment between two sequences of observations. It finds the alignment by warping the time dimension with certain constraints. Hence, DTW can suitably aid in Time series classification methods.

In this paper, we propose a combination of k-NN and DTW to classify the Sleep data. The training data (labeled) set consists of 70% of sleep data (120960 seconds), the validation set consists of 15% of sleep data(25920 seconds) and the Test data set consists of the rest 15% of sleep data (25920 seconds). As the sample rate used is 250Hz, each second consists of 250 sample values. The k-NN algorithm establishes a Euclidean space consisting of labeled data (training set). The DTW Algorithm receives the query observation and computes the similarity of the query observation with each labeled observation. On the basis of similarity computation from the DTW, the query observation is allocated a location on the Euclidean space. The user defines the 'k' value and hence we choose an odd number (here, we choose 7 as the 'k' value) in order avoid a conflict arising due to the tie in the majority voting (i.e. both labels have equal vote, so the label of the query can't be decided). Therefore, a query observation (each test second) is classified as non-movement ('0') or movement ('1') accordingly in Binary Classification task. The test data in classified in batches of 30 seconds epochs and the Machine learning model outputs 30 corresponding labels (i.e. labeled for each second).

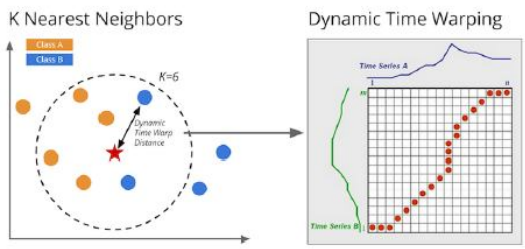

Figure 2: The k-NN and DTW Classifier

IV. EXPERIMENTAL RESULTS

The Validation set was used to tune the hyper-parameters. It consists of 25920 seconds (15% of

data). The test data set consists of 25920 seconds (15% of data available) from various subjects. The data is classified in batches of 30-second epochs. The ground truth labels for the data set are created by labeling the data by verifying with the video capture of the subjects (labeled '0' or '1' for each second). The performance of the Classifier was evaluated on the basis of confusion matrix as well as classification accuracy. The confusion matrix depicts a low percent of false-negatives and false-positives. After tuning the hyper-parameters, the validation accuracy came out to be 98.72%. The test set gave an accuracy of 98.17%.

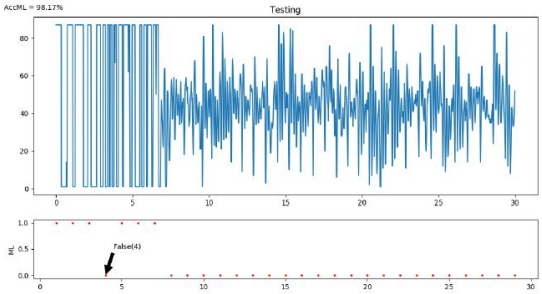

Figure 3: Testing and Evaluation

	Predicted(Non-Movement)	Predicted(Movement)
Actual(Non-Movement)	16654	370
Actual(Movement)	96	8910

Table 1: Confusion Matrix

CONCLUSION

In this paper, we proposed a robust and powerful classifier to classify movement from non-movement patterns occurring during sleep. We further utilize the labels to classify the movement into four categories i.e. no movement, light movement which lasts for 3 seconds, heavy movement which lasts maximum for 6 seconds and very heavy movement which lasts for more than 6 seconds. The classification report helps to infer restlessness, quality of sleep etc. One of the major drawbacks of the classifier is the fact that it is computationally expensive as it has a time complexity of $O(n^2)$. However, by constraining the warping window one can attain results comparatively faster with the trade-off with accuracy to a small extent and other enhancements include dynamic warping window [10]. There are many improvements that could be made to the k-NN and DTW (exploring parallel processing). All in all, the k- Nearest Neighbors and Dynamic Time Warping Classifier yields accurate results in classifying the sleep movements.

REFERENCES

[1] Harvey AG, Stinson K, Whitaker KL, Moskovitz D, Virk H. The Subjective Meaning of Sleep Quality: A Comparison of Individuals with and without Insomnia. *Sleep*. 2008;31(3):383-393.

[2] Metsis, Vangelis & Galatas, Georgios & Papangelis, Alexandros & Kosmopoulos, Dimitris & Makedon, Fillia. (2011). Recognition of sleep patterns using a bed pressure mat. ACM International Conference Proceeding Series. 9. 10.1145/2141622.2141633.

[3] Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. 2017. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Discov*. 31, 3 (May 2017), 606-660. DOI: https://doi.org/10.1007/s10618-016-0483-9

[4] Tal A, Shinar Z, Shaki D, Codish S, Goldbart A. Validation of Contact-Free Sleep Monitoring Device with Comparison to Polysomnography. *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine*. 2017;13(3):517-522. doi:10.5664/jcsm.6514.

[5] Ni H., Zhao T., Zhou X., Wang Z., Chen L., Yang J. (2015) Analyzing Sleep Stages in Home Environment Based on Ballistocardiography. In: Yin X., Ho K., Zeng D., Aickelin U., Zhou R., Wang H. (eds) Health Information Science. HIS 2015. Lecture Notes in Computer Science, vol 9085. Springer, Cham

[6] Yakowitz, S. (1987), NEAREST-NEIGHBOUR METHODS FOR TIME SERIES ANALYSIS. Journal of Time Series Analysis, 8: 235–247. doi:10.1111/j.1467-9892.1987.tb00435.x

[7] T. Cover and P. Hart. Nearest neighbor pattern classification. Information Theory, IEEE Transactions on, 13(1):21–27, 2002

[8] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In AAAI-94 workshop on knowledge discovery in databases, pages 229–248, 1994.

[9] A. E. Flores *et al*., "Pattern Recognition of Sleep in Rodents Using Piezoelectric Signals Generated by Gross Body Movements," in *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, pp. 225-233, Feb. 2007..

[10] B. Ben Ali, Y. Masmoudi and S. Dhouib, "Accurate and fast Dynamic Time Warping approximation using upper bounds," *2015 38th International Conference on Telecommunications and Signal Processing (TSP)*, Prague, 2015, pp. 1-6.

★★★