Improving Stress Forecasting using LSTM Neural Networks

Terumi Umematsu, Akane Sano, Sara Taylor, and Rosalind Picard

Abstract—Accurately forecasting stress might enable people to make changes that could improve their future health. In this paper, we examine how accurately previous N-day multi-modal data from wearable sensors, mobile phones and surveys can predict tomorrow’s level of stress using long short-term memory neural network models (LSTM), logistic regression (LR), and support vector machine (SVM). Using 1231 days of data from 201 college students, we find the LSTM outperforms the LR and SVM with the best results reaching 83.3–84.1% using 5–7 days of prior data.

I. INTRODUCTION

Accurately forecasting well-being (i.e., stress, mood and health levels) has a number of important clinical benefits and could enable better self-management of an individual’s behavioral choices in ways that might prevent damage to their physical and mental health. For example, stress increases susceptibility to infection and illness [1], and knowing that personal behaviors are leading to increasing stress might motivate somebody to make changes to their schedule, such as cancelling a commitment or getting more sleep, in order to prevent excessive stress tomorrow night.

Previous work has shown that tomorrow’s mood, stress and health levels can be predicted with 78-82% accuracy based on today's physiological and behavioral data [2]. In addition, it is said that human mental condition is affected by the past few days [1]. Suhara and colleagues have shown that depressed mood can be predicted using survey data of the past several days [3]. Specifically, depressed mood was predicted from 14 days of survey data with an AUC score of 0.886. However, it has not been examined whether well-being prediction accuracy improves using several days of data from other modalities. We compare the stress prediction accuracies from static (SVM, LR) and time-series (LSTM) models using data using the previous 1 to 7 days of behavioral survey, physiological, phone and mobility data.

II. EXPERIMENTS

The data in this experiment were collected by a study to measure Sleep, Networks, Affect, Performance, Stress, and Health using Objective Techniques (SNAPSHOT) [4]. 201 college students from 5 cohorts were monitored for 30 days each. The study gathered multi-modal data, including daily smartphone, physiological, and behavioral data. Stress scores were collected every evening using self-reported values from 0-100. The days were split into groups of the 40% highest-stress and 40% lowest-stress scores to create binary labels to classify high/low stress. We used 1231 days of data and computed 380 daily features including 42 behavioral survey features (excluding self-reported stress scores), 173 physiology features, 150 smartphone features and 15 mobility features. We compared the accuracy of forecasting tomorrow night’s high/low stress for each modality (survey, physiology, smartphone and mobility) using an SVM classifier with a radial basis function kernel, LR, and LSTM. In order to predict tomorrow night’s stress levels, we varied the input data from today (1 day) to the past one week (7 days) for the LSTM models and evaluated the model performance using 5-fold cross-validation. The LSTM was trained using RMSprop with binary cross-entropy loss, a sigmoid activation function, and an iteration number of 1000.

III. RESULTS

The stress prediction accuracy using SVM (1 or 7 days), LR (1 or 7 days) and LSTM (1, 2, 3, 4, 5, 6, or 7 days) of data is shown in Fig. 1. The baseline accuracy is 50.7%. The results show that accuracy is improved by using the LSTM instead of static models and by using a longer history in each modality (one way ANOVA, Tukey’s HSD test, p<0.05). Particularly, we see that the best results use data from at least the previous 4 days. In addition, we confirmed that using time-series models improves the prediction of stress when using survey data, physiology data, or phone data. The best results overall of 84.1% (AUC 0.837) were obtained using all the features with 5 days of data and the best single modality results obtained 83.2% (AUC 0.829) using physiological features and 7 days of data. The improvement of the time-series model over the static models was largest for the physiology and mobility features. In future work, we plan to examine other LSTM structures and evaluate forecasting stress levels further in the future (i.e., a week from today). In addition, we will consider adaptive methods to fill in missing data with time series information.

Figure 1. Stress prediction accuracy for 2 support vector machine (SVM), 2 logistic regression (LR), and 7 long short-term memory neural network models (LSTM)

REFERENCES