A Medication Compliance Monitoring System using Machine Learning and Distributed Computing

UNIVERSITY OF SAN FRANCISCO

Hai Vu Le, Katja Wittfoth, Donya Fozoonmayeh & Diane Woodbridge

INTRODUCTION

This research aims to detect medication-taking behaviors based on accelerometer and gyroscope data collected by smartwatches. The end goal is to develop a smart machine learning application that could provide patients with accurate and timely reminders to improve their medication adherence.

To tackle the challenge of processing massive high-frequency sensor data in real time, we take a distributed computing approach with an end-to-end pipeline built on distributed systems.

BACKGROUND

The low medication-adherence rate in the US is reported to be between 25% to 50% in 2010, resulting in an annual loss between $100 and $300 billion which accounts for 3% to 10% of total healthcare costs in the country. Also, the social and economic impacts on patients' and their family are apparent.

Traditional efforts to enhance adherence rates including log sheets or text message reminders do not yield high effectiveness for requiring high involvement of patients. Technology-empowered methods promise better solutions as they address the major issue of poor adherence: patients forget to take medication.

The increasing popularity and highly positive social acceptance that smartwatches and likewise wearables make them the ideal devices for a medication-adherence enhancing application.

DATA ACQUISITION AND PROCESSING

We collected three-dimensional sensor data of activities including medication intake (pill and liquid, using the dominant and non-dominant hand), texting, writing, drinking bottled water, and walking among 24 study subjects using LG Watch Sport devices. The time series record for an activity ranges between 2.85 to 16.13 seconds.

To account for the variation in activity duration and the difference in subjects' pace, all records were discretized into the same number of time windows. Then for each dimension, we extract 18 descriptive statistics out of each window, including mean, maximum, minimum, and quantiles as features to train machine learning models. The choice of number of time windows are 5, 10, 20, 30, 40, 50, and 60.

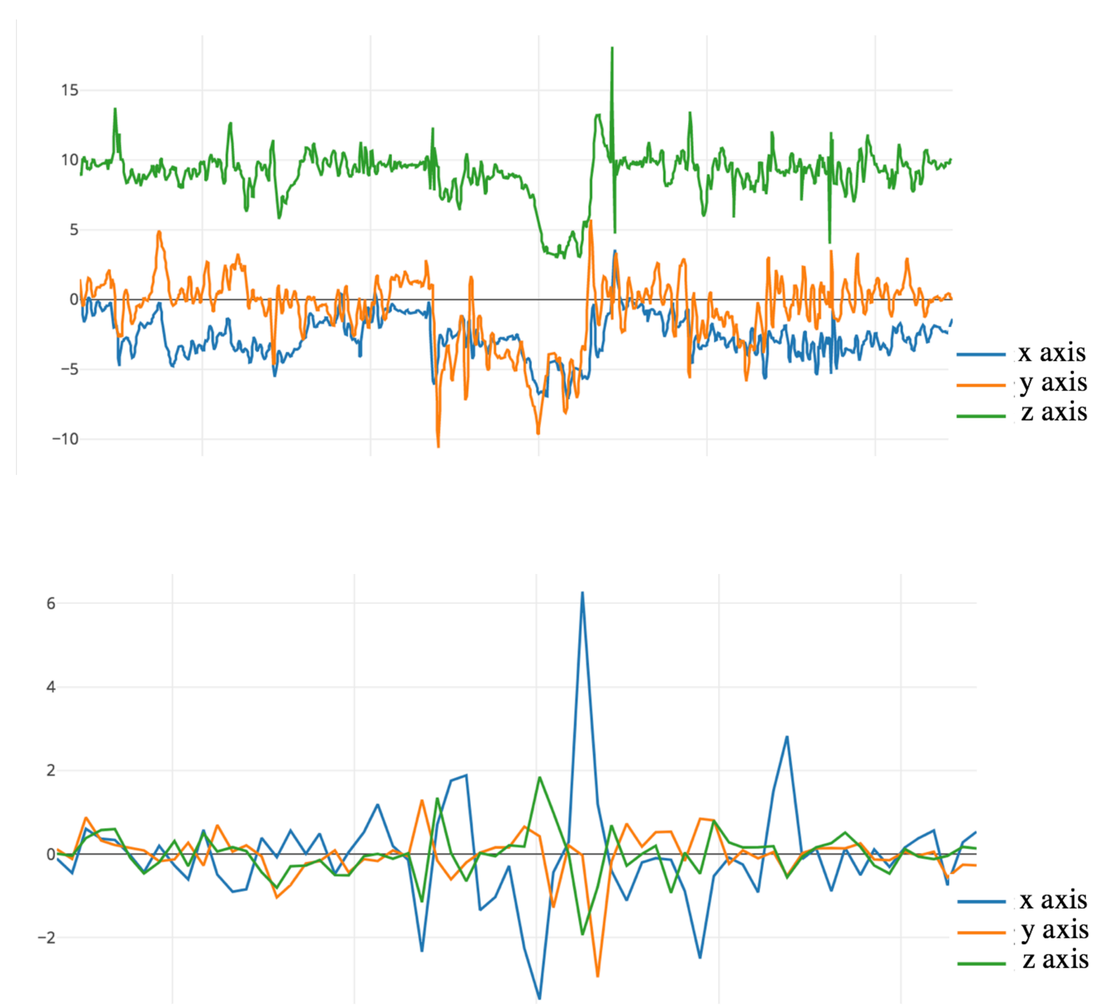

Figure 1: Accelerometer and gyroscope data stream during medication intake

DATA PIPELINE

The massive nature of high-frequency sensor data coming real-time from multiple patients together with the large number of features required to train high-performing machine learning models necessitate large storage and computing capability.

Our solution is to send raw data from smartwatches to Amazon Web Service S3 for storage, use Spark for data processing, store features in a MongoDB database system, and then use Spark for machine learning modeling and prediction.

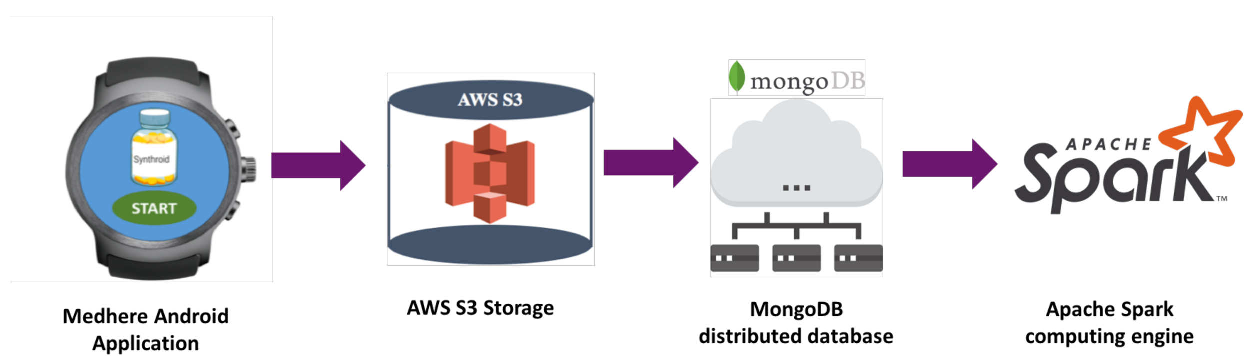

Figure 2: The end-to-end pipeline built on distributed systems

MACHINE LEARNING ALGORITHMS

We trained various machine learning models for this classification problem, including Logistic Regression, Random Forest, and XGBoost - both computationally inexpensive models and those with high computational cost yet empirically proven strong performance as both accuracy and speed are important for this problem, given that prediction time is especially important to provide timely reminders to patients. Additionally, to perform well on speed, we use Apache Spark – a distributed computing engine for modeling and prediction.

EXPERIMENT RESULTS

We used precision, recall, and execution time as metrics to evaluate our models. Random Forest (RF) algorithm performs the best for this task of identifying medication-taking behavior.

Our best RF model yields a precision score of 0.98 with data discretized into 40 windows. It also outperformed other models in terms of speed. The training time is 14 seconds and the prediction time is 0.17 seconds for 24 study subjects.

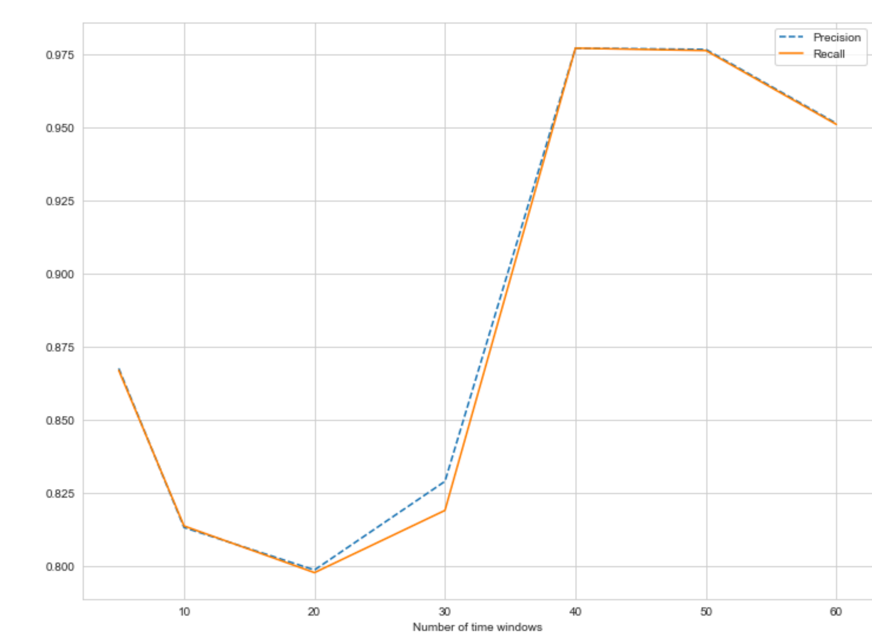

Figure 3: Precision and Recall of a Random Forest model

CONCLUSION

This study shows a high potential of using smartwatch sensor data on a distributed system as the path forward to enhance medication adherence. In future studies, we will utilize more sensor readings including those from sound and near-field communication (NFC) sensors. Regarding modeling, we will apply neural network models to explore the applicability of high-performing deep learning techniques to this problem.