Using SVMs to Classify Activities
Proposal

Tom Stuckey

22 Sept 2020

Idea

This research will examine the efficacy of support vector machines (SVMs) in classifying activities of people based on passive radio frequency identifiers (RFIDs). This multi-class classification problem will use R with several processing and visualization libraries to support the analysis. Hereafter some additional high-level support information is presented.

Anticipated Challenges

Several challenges are anticipated. Data is likely to need to be cleaned where empty values are either removed or replaced with an appropriate representative replacement. A cross-validation strategy will need to be implemented and developed to improve the strength of the findings. It is likely strategies to both identify and handle both noise and anomalies will need to developed in accordance with best practices [1].

Potential Results

While the overall goal is to develop an efficient classifier, there are several other potential results. First, although unlikely, a new class of human movements could present from an analysis of the data. Second, substantial cross-correlation could represent in the data such that certain features should either be omitted or combined. Third, analysis might reveal certain RFID sensor configurations are more effective at identifying movements than others.

Data Sources

The machine learning repository archive from University of California is an excellent source of interesting data. Health researchers from Australia donated a set of motion data from older adults as measured from wireless RFID sensors [2]. Principally, the data appears to leverage frontal, vertical, and lateral axis readings across a time dimension to classify activities as: 1. Sitting on bed, 2. Sitting on chair, 3. Lying down, or 4. Walking. Additionally, there is supporting information like signal strength and gender of the participant.

Algorithm

SVMs will be leveraged to perform this multi-classification analysis. SVMs fundamentally optimize the soft margin containing support vectors from different classes; they are popular and are typically employed in binary classification problems [3]. Accordingly, this analysis will more likely leverage a one-vs-all approach to examine membership in a given activity class vs. not in a given activity class rather than a one-vs-one (pairwise combination) approach.

For the implementation, R will be leveraged. Additionally, the tidyverse package for data processing and visualization will be used. Liquid SVM package will be employed for actual SVM implementation engine [4].

References

[1] Aggarawal, C. A. (2016). Outlier Analysis (Second Edition ed.). http://charuaggarwal.net/outlierbook.pdf

[2] Torres, R. L. S. T., Ranasinghe, D. R., & Visvanathan, R. V. (2016). Activity recognition with healthy older people using a batteryless wearable sensor Data Set. UCI Machine Learning. https://archive.ics.uci.edu/ml/datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor

[3] Sarker, I.H., Kayes, A.S.M. & Watters, P. Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. J Big Data 6, 57 (2019). https://doi.org/10.1186/s40537-019-0219-y

[4] liquidSVM (1.2.4). (2019). [SVM package]. CRAN. https://www.r-pkg.org/pkg/liquidSVM