Predicting the type of hot beverage based on coffee machine vibrations

TARTU ÜLIKOOL
arvutiteaduse instituut

Andreas Vija
Rasmus Maide

IT Akadeemia
toetab Skype™

Introduction

All across the world, motors power billions of machines. Most motors also produce vibration. These vibrations are different for different processes, usage patterns and stages of motor health. Knowing how to “listen to” these vibrations well enough would allow for many feats: predicting machine failure early, predicting harmful usage patterns, learning machine work patterns and many more. As such, immense benefit could be gained in all areas of industry from knowing how to listen to machine vibrations. Our goal is to analyze the vibrations of one coffee machine and try to predict which beverage is being made. This is the first proof-of-concept step towards gaining more insight into machine vibrations.

Data

Our data was private and provided by the company Proekspert. In this project we viewed a subset consisting of one coffee machine’s timeseries data collected over 2.5 months with around 200 measurements per second. For each measurement, data contained 6 different values of vibration ('Xaccel', 'Yaccel', 'Zaccel', 'Xgyro', 'Ygyro', 'Zgyro'). In addition, there were 1646 labelled instances composed of a time and the beverage made at that moment. The initial total data size was around 100 GB.

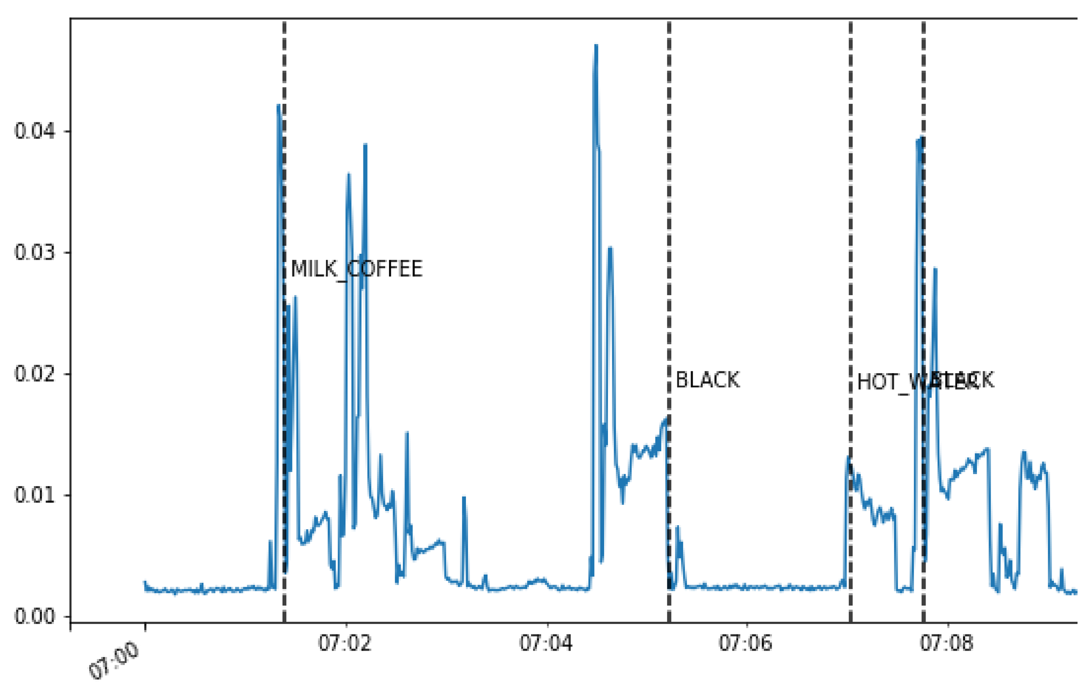

Figure 2. Different vibration values over 15 minutes and drinks ordered during that time, after initial preprocessing

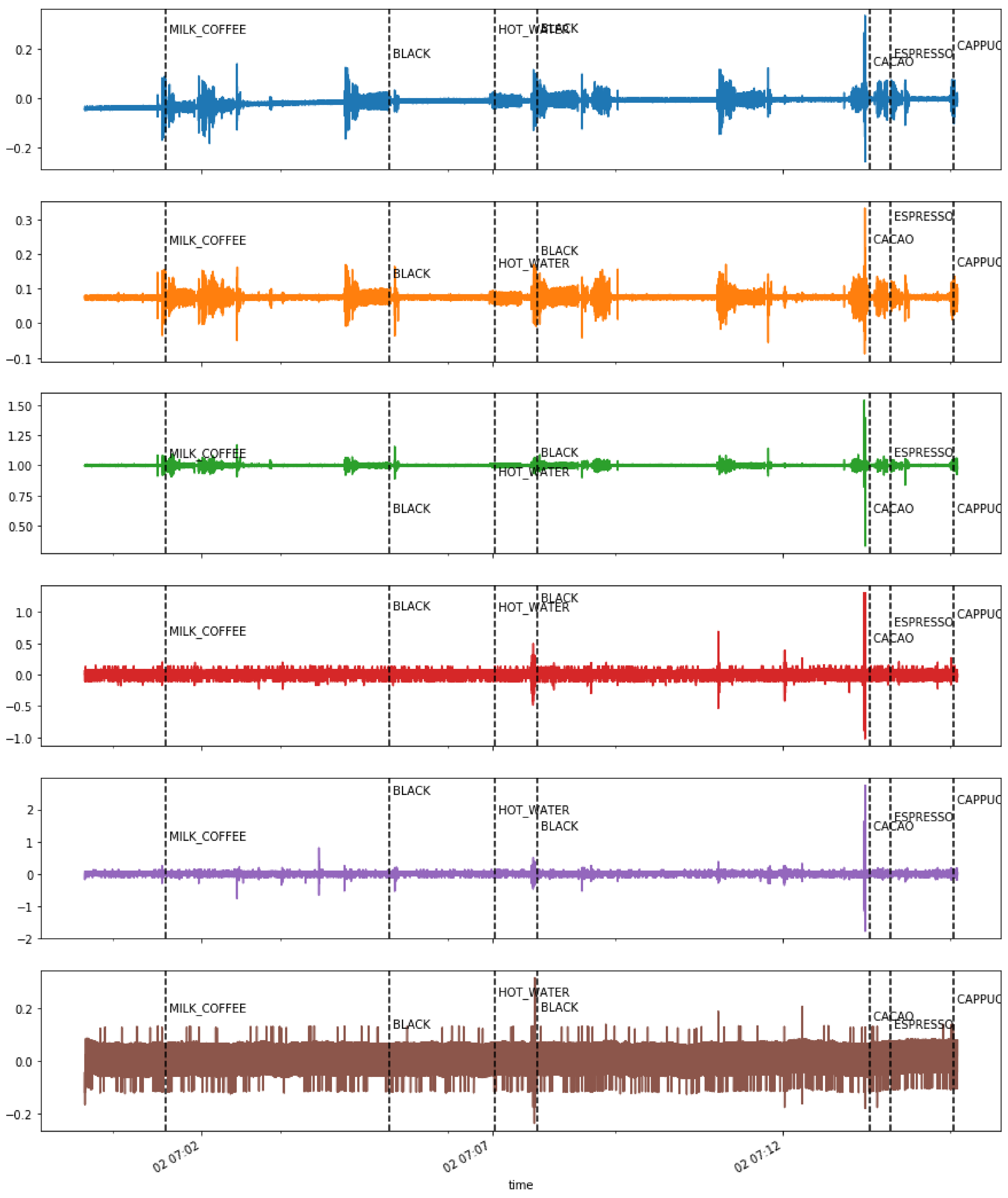

Figure 1. Initial visualization of each parameter

Preprocessing

As the initial visualization showed, the first three values are largely similar and the last three very chaotic. Therefore, we worked with just the 'Xaccel' value. We centered the data by 1 minute snippets, saved only the median amplitude of the vibration for each second and levelled off the biggest vibrations. Then for each label we defined the event as the activity surrounded on both sides by at least 4 seconds of low vibration. To unify events of different lengths for machine learbing, we finally saved for each labelled point the means of each 1/12 of the data, getting for each label 12 features describing the event.

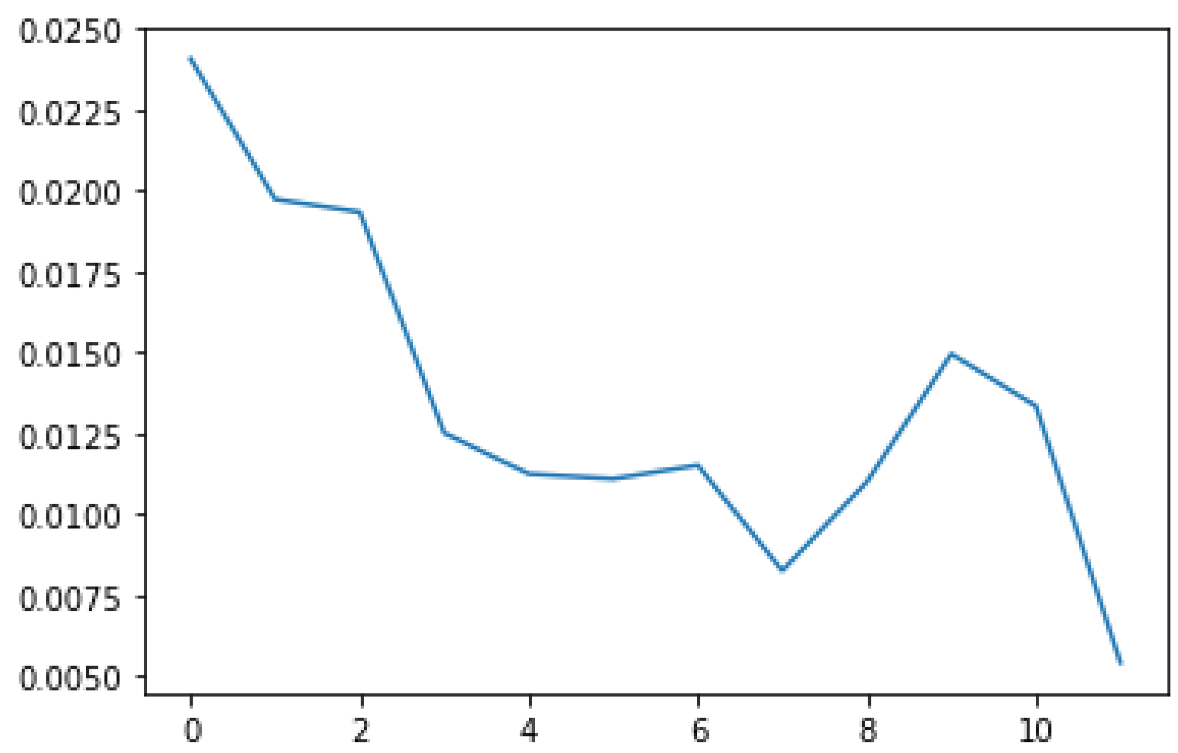

Figure 3. One beverage making event in the final form

Machine Learning

The dataset was very unbalanced, with black coffee represented 509 times while there were just 14 instances of mocca. However, balancing the dataset did not improve the results because the test set would also have to be unbalanced. To get better results with machine learning, we used standardization of the data. We obtained the best results by using random forest classifier with at least 70 trees and using Gini impurity. This yielded an accuracy of over 72% on the test set (using 5-fold cross-validation). Since a lot of the errors were caused by similar drinks in the data, such as black coffee and double black coffee, we also trained a model to classify the five most popular drinks. This got an accuracy score of over 86%.

	All drinks	Top 5
SVM	70%	84%
KNN	70%	83%
Random forest	72%	86%

Figure 4. The accuracies of different models with the best found parameters

Results

- We misjudged the data somewhat in the initial analysis
- Data quality was good, but labels were at odd times which made event extraction hard and very imprecise.
- Best model was random forest, which achieved 72% accuracy on all data and 86% accuracy on top 5 beverages
- Even though we did not reach the initial goal of 90%, we can be satisfied with how we managed to classify beverages. There were too many similar drinks, overlapping data and bad label times to get the results we initially wanted.

Contact:
Andreas Vija
andreasvija@hotmail.com
Rasmus Maide
rasmusmaide@gmail.com

Source code:
https://github.com/andreasvija/coffee-ekspert