

Motivation

Francisco Vargas

February 14, 2016

1 Motivation

1.1 Practical Motivation

This project was originally supposed to solve a real world problem for a local Edinburgh based company called IceRobotics.ltd .

IceRobotics domain is in the livestock industry. Their main product is called the iceQube which is a 3 axes 6hz sample accelerometer that is attached to cattle in farm in order to monitor and manage cattle behaviour.

The company provides a set of basic statistical algorithms that attempt to detect health states of the animal such as oestrus or lameness. Since cattle are naturally animals of habit with recurring behavioural patterns at a daily and weekly basis thus a good postulate to tackle this classification problem seemed to view the data in the frequency domain and use these representations as a features for machine learning based algorithms.

1.1.1 Specific Tasks

From a practical viewpoint I am interested in tackling the classification process in both a economically efficient and high performing solution. With high performing solution I mean that I wish to fit the model over a domain which represents the underlying dynamics (nature of data) correctly both from a theoretical modelling perspective and a good set of results in terms of accuracy, precision, recall , confidence and other metrics that asses good performance in an inference (learning) based method. Thus two very relevant seeming tasks that I will study are:

1.1.2 Assessing Number of Devices to Use

One question that one may ask. Is how many devices to use on each animal (for example one on each leg, total of 4). This question can be answered using methods from cooperative game theory.

In our scenario we interpret each device as a player (agent) and ask which co-operation are meaningful among the players available in the game (coalitional game). We can formally define such scenario in the following manner :

1. *Accelerometers* = $A_c = \{1, \dots, n\}$, (agents);
2. Coalitions: $\forall C \subset \mathbb{P}(A_c)$, s.t. $(\mathbb{P}(A_c) \equiv \mathbf{2}^{A_c})$;
3. $G = A_c$ is the grand coalition;
4. $\nu : \mathbb{P}(A_c) \rightarrow \mathbb{R}$ is the pay-off (utility) function of the game s.t. $\nu(\emptyset) = 0$ (ν could be the accuracy when classifying the test set) ;
5. a cooperative game is the pair $\Gamma = (A_c, \nu)$;

Having set up such game mechanism we can ask questions such as:

1. Which accelerometer placement contributes the most towards our modelling;
2. Which coalitions (combinations of accelerometers) play the best together in our modelling task;

1.1.3 Spectral Analysis

As we mentioned before the periodicity underlying the dynamics of the of cattle motion make the frequency space a good domain in which to represent our data.

Due to the non linearity that some of the useful representations in this frequency space posses it becomes hard for a machine learning algorithm to learn these representations under real constraints such as a limited dataset.

Thus some of these transformations will be embedded in the models and allowed to be relaxed when optimizing the likelihood or posterior of the classifier.

1.1.4 Time Series Models

1.2 Physical Motivation

The task can be summarized as learning a model which allows us to accurately and confidently classify the state of the time varying process outlined by the motion of an individual animal.

1.2.1 Fourier Transform

We can formally describe motion of the animal as a signal $\vec{x} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ where n may be the number of dimensions we are measuring in (three) times the different time of motion metrics (acceleration, magnetic field, change in orientation, and such).

As we described before \vec{x} exhibits periodic behaviour that both characterizes its state and evolves given the state of animal changes. Due to the periodicity of \vec{x} it may be tempting to assume that the signal is a superposition of sinusoids with a defined period T , which is known as a Fourier series:

$$\vec{x}_j(t) = \sum_{n=-\infty}^{\infty} C_n e^{\frac{in\pi x}{T}}, \quad \text{s.t. } C_n \in \mathbb{C} \wedge \vec{x}_j \in \mathbb{R}$$

Where C_n are orthonormal coefficients defined by:

$$C_n = \int_{-T}^T \frac{dt}{2T} \vec{x}_j(t) \cdot e^{\frac{-in\pi x}{T}}$$

The problem with this representation of the frequency domain is that it assumes that $\vec{x}_j(t)$ repeats itself in an identical manner outside the range $[-T, T]$ which is not the case since as we discussed above the period of the animals in study evolves in time thus we need a better representation. Taking the limit $T \rightarrow \infty$ under the reasoning that a function with no well defined period is a function whose period goes to infinity:

$$C_n = \lim_{2T \rightarrow \infty} \left\{ \int_{-T}^T \frac{dt}{2T} \vec{x}_j(t) \cdot e^{\frac{-in\pi x}{T}} \right\}$$

The limit will make $\frac{-in\pi x}{T}$ become a continuous variable namely ω . Labelling $2T$ as the period T' we have:

$$C_n = \lim_{T' \rightarrow \infty} \left\{ \int_{-\frac{T'}{2}}^{\frac{T'}{2}} \frac{dt}{T'} \vec{x}_j(t) \cdot e^{\frac{-2in\pi x}{T'}} \right\}$$

Multiplying both sides by T' we arrive to a transformation known as the Fourier transform:

$$\begin{aligned} \mathbf{FT}\{\vec{x}_j(t)\} &\equiv \vec{X}_j(\omega) = C_n \cdot T' = \lim_{T' \rightarrow \infty} \left\{ \int_{-\frac{T'}{2}}^{\frac{T'}{2}} dt \vec{x}_j(t) \cdot e^{-ikx} \right\} \\ \vec{X}_j(\omega) &= \int_{-\infty}^{\infty} dt \vec{x}_j(t) \cdot e^{-ikx} \end{aligned}$$

This yields a much better representation in the frequency domain for our given problem than the original Fourier series from which the limit was taken. Whilst this captures that the function does more than copying itself outside a defined range it does not capture that the frequency space itself evolves through time.

1.2.2 Short Time Fourier Transform

The short time Fourier transform is a Fourier related transformation which is carried out by transforming only short windows of the signal:

$$X(\tau, \omega) = \int_{-\infty}^{\infty} dt \Pi_{l/2}(t - \tau - l/2) \vec{x}_j(t) e^{i\omega t}$$

Where $\Pi_{l/2}(t)$ is the top hat function centered at 0 with width l .

The issue of taking out rectangular windows of our signal (multiplying by the top hat function) is that in Fourier space this equates to convolution with the sinc function (convolution theorem):

$$\mathbf{FT}\{\Pi_{l/2}(t) \cdot \vec{x}_j(t)\} = \frac{1}{2\pi} \mathbf{FT}\{\Pi_{l/2}(t)\} * \mathbf{FT}\{\vec{x}_j(t)\}$$

Where $*$ denotes convolution.

$$\mathbf{FT}\{\Pi_{l/2}(t)\} = \int_{-\infty}^{\infty} dt \Pi_{l/2}(t) e^{-i\omega t}$$

$$\mathbf{FT}\{\Pi_{l/2}(t)\} = \int_{-l/2}^{l/2} dt e^{-i\omega t}$$

$$\mathbf{FT}\{\Pi_{l/2}(t)\} = - \left. \frac{e^{-i\omega t}}{i\omega} \right|_{t=-l/2}^{t=l/2}$$

$$\mathbf{FT}\{\Pi_{l/2}(t)\} = \frac{e^{i\omega l/2}}{i\omega} - \frac{e^{-i\omega l/2}}{i\omega} = l \frac{\sin(l/2\omega)}{l/2\omega} = l \cdot \text{sinc}(l/2\omega)$$

Thus:

$$\mathbf{FT}\{\Pi_{l/2}(t) \cdot \vec{x}_j(t)\} = \frac{l}{2\pi} \text{sinc}(l/2\omega) * \mathbf{FT}\{\vec{x}_j(t)\}$$

Effectively we are blurring the Fourier transform of our motion signal with a sinc function which heavily distorts it. This is also known as the uncertainty principle in signal processing and it represents the trade-off between measuring accurately in the time domain vs measuring accurately in the frequency domain just like Heisenberg's uncertainty principle which models the same over momentum and space.

In order to solvent this issue we multiply our top hat function with a smoothing function such as a Gaussian for example which equates to having a window function W that is 0 outside a specified range yielding the general for of the short time fourier transform:

$$\mathbf{STFT}\{\vec{x}_j(t)\} = X(\tau, \omega) = \int_{-\infty}^{\infty} dt W(t + a - \tau) \vec{x}_j(t + a) e^{i\omega t}$$

With an appropriate choice for the window length of W and the overlap term a this may yield a good spectral representation of our signal in which evolution of the spectrum is taken in to account (thus modelling the model signal as a non-stationary process). The magnitude of this result is referred to as the spectrogram which is a tool that will be recursively used through this project.

1.2.3 Discrete Time

Talk about shah function and the discrete equivalents of the previous two transforms.

1.2.4 Probabilistic Approach

Motivate the idea of a spectral prior as a regularizer of some sort.