Can a CNN identify abnormality in advertising dynamics with wavelet transformed images?

1. Introduction

In neural science and medical engineering, EEG signal analysis is processed with EEG Data which is measured through electrodes attached to the scalp. The spontaneous electrical signals generated following the neuronal activities can be acquired in a noninvasive manner over time. Since each channel of electrodes typically shows a certain pattern of time series dynamics, have been able to discover omen symptoms of epilepsy, infarction and other diseases by anomaly detection methods.

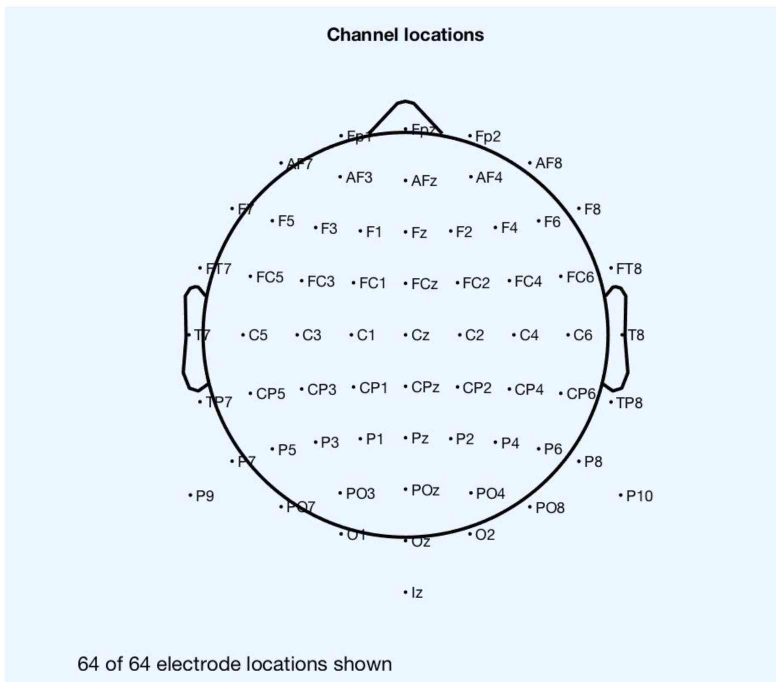


Fig 1. Electrodes channels to measure EEG signal

As neural scientists found patterns in neuronal activity and its EEG data, we also could find certain patterns shown in the digital marketing data not only in periodically but also in categorically. Through three experiments, it is found that AlexNet can identify abnormalities in advertising click dynamics with wavelet transformed input images. The detection accuracy was 0.857 in the experiments.

2. Theoretical Background

Wavelet transformation is an advanced fourier transformation which acknowledges and takes the weight of time into account when calculating convolution operator. In Fourier transform, the kernel is sine wave for the convolution operator that is in real 'sum of dot product between two vectors'. The kernel of Fourier transform can be defined as following equation

$$y(t) = Asin(2\pi ft + \theta),$$

where A is the amplitude which affects the height of the sine wave, f is the frequency, the distance between one peak to the next peak, and θ is phase angle. Thus, if θ is equal to π / 2, the $sin(t + \theta)$ is also equal to $cos(t)$.

However, derived from Euler's formula using complex sine wave signals can be presented on the complex plane.

$$Me^{ik} = M(cos(\theta) + isin(\theta)).$$

During the implementation of complex sine wave, parameter i and k need to be carefully set up since these parameters determine the important characters of signals. They are originally from Gaussian curve which is a characteristic symmetric 'bell curve' shape where the parameter a is the height of the curve's peak. Parameter b is the position of the center of the peak and c presents the standard deviation, sometimes called the Gaussian RMS width) controls the width of the 'bell'.

$$f(x) = ae^{-(x-b)^2/(2c^2)},$$

$$e^{-(t-peaktime)^2/(2s^2)}.$$

Thus, parameter i represents the position of the center peak and k presents the width of the bell curve respectively. The advantages that we can take from the complex sine wave are 1) it enables transformation between cartesian coordinates and polar coordinates which guarantee us to be able to find the vector of phases which we will use in the experiment (Fig 7) and 2) it enables to represent vectors on complex plane, composed of real Axis and imaginary Axis. Through presenting a vector as a pair of complex numbers, we can derive tremendously important solutions in many fields.

Pertaining to the proof which is still important please refer this article. Corresponding to the article, we can conclude these equations

$$real\{e^{j\omega t}\} = cos(\omega t),$$

$$imaginary\{e^{j\omega t}\} = sin(\omega t).$$

While the complex sine wave is still stationary which does not take the weight of time into account, wavelets are non-stationary and take the weight of time into account. Fig 2 shows the simulated wave of morlet wavelet transform.

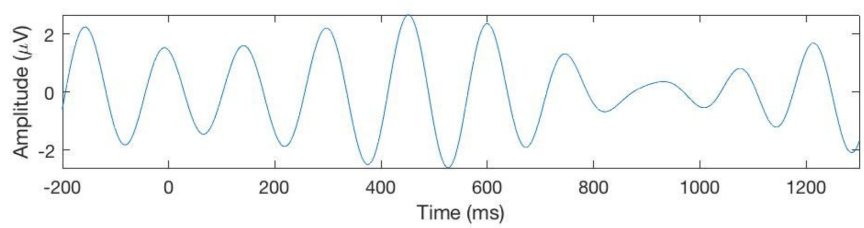


Fig 2. simulation of morlet wavelet kernel

To generate wavelet images, in the experiments three wavelets [morse, gabor, bump] are used.

3. Methods and Materials

Data

The original data utilized in the experiments are from Criteo Lab which can be downloaded here. Also data preprocessing was proceeded by python which can also downloaded here. (Loading entire data (1.5GB Text file) causes memory error. Please split the raw data first.)

Feature Extraction

As explained in section 2, there are three types of data that we can produce with wavelets and thus, combination of extracted image could vary. In the experiment, mainly two types of information, 1)magnitude and 2)angle phase, are utilized to generate training images and test image sets.

Classification - Anomaly Detection

With the first three weeks of click data (two week for train, one week for test) which has the time unit as 5 mins or 1 min, proposed algorithm has been applied. Fig 3 and 4 shows the normalized click and conversion data. Correlation coefficient between the two is high (on average over 0.8) and abnormal patterns are shown in the same dates.

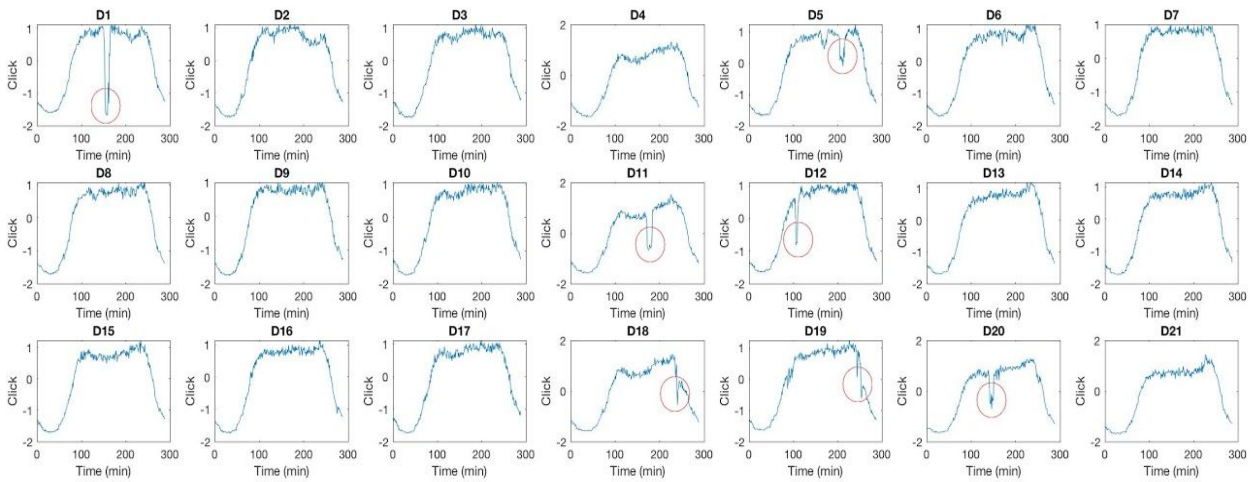


Fig 3. Normalized daily click highlighted with a circle which represents the abnormality

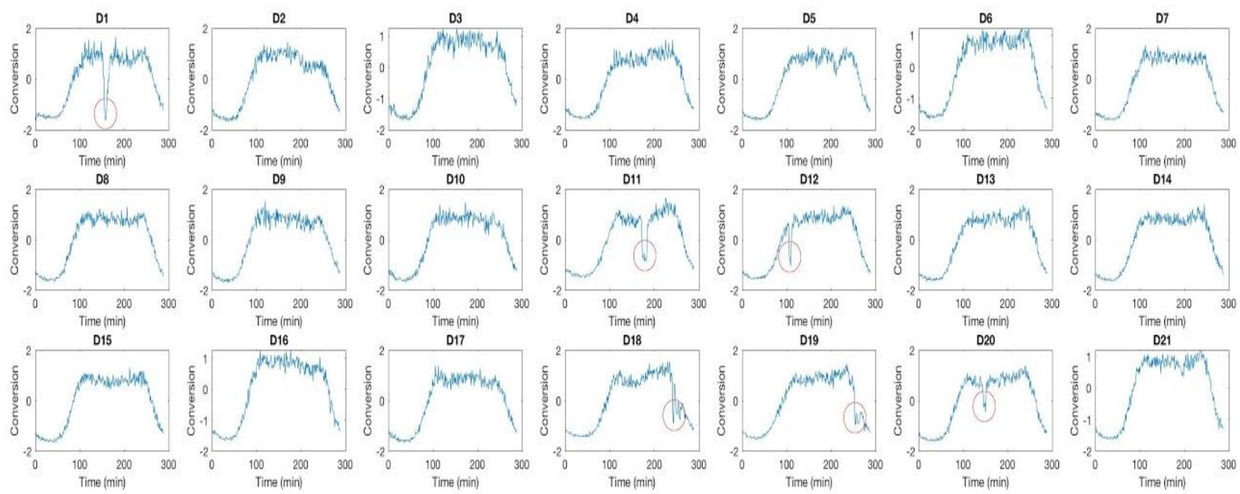


Fig 4. Normalized daily conversion highlighted with a circle which represents the abnormality

To generate input images for training and test images for performance evaluation, three types of wavelet transformation (Morse, Gabor, and bump wavelet) have been chosen. (All the work was coded in matlab which can be downloaded here.) Fig 5 to 7 show the result images that those three wavelets transform the daily click and daily conversion data with different time units (1minute and 5 minutes). Fig 5 and 6 are the magnitude, length of the position vector, and Fig 7 is the phase angle, $arctangent(y/x)$. The sharp decrease in Day1 in Fig 3 and 4 is more clearly discovered in each wavlet's images. As it comes out in Fig 5 and 6, the larger time unit makes the sharp drop more noticeable in the images with more noticeable noises. In the case of phase angle, a set of vertical stripes is found at the position where the sharp drop occurs.

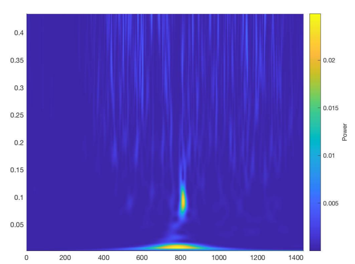
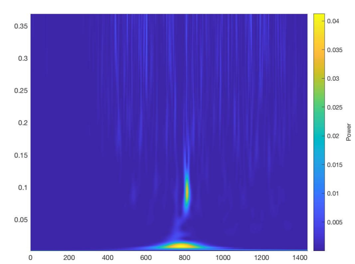
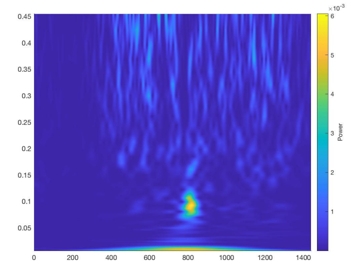
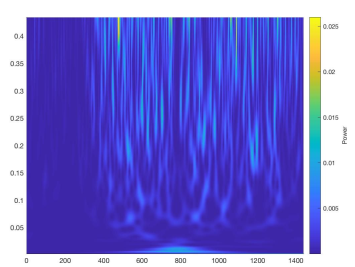
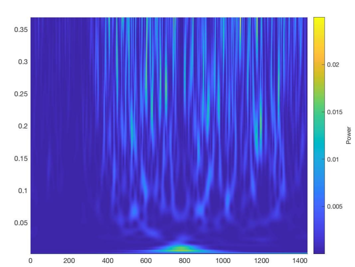
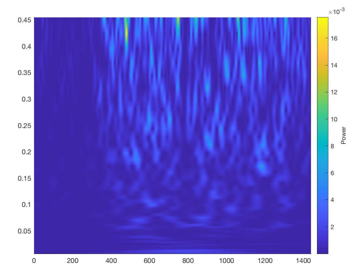
		
Morse Click D1 (1min)	Gabor Click D1 (1min)	Bump Click D1 (1min)
		
Morse Conversion D1 (1min)	Gabor Conversion D1 (1min)	Bump Conversion D1 (1min)

Fig 5. Morse, Gabor, and bump wavelet magnitude with abnormality in D1 where the time unit is 1min

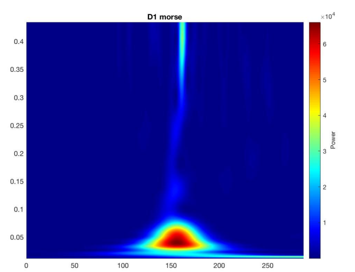
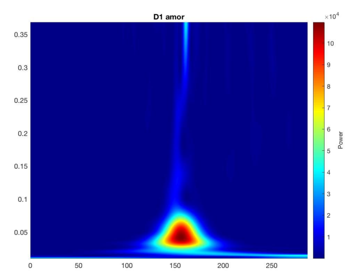
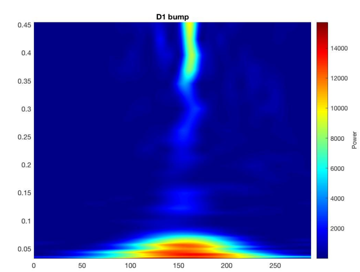
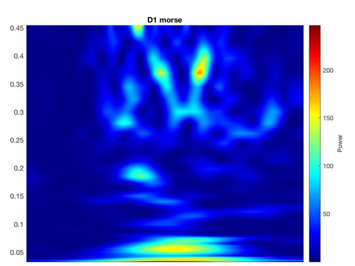
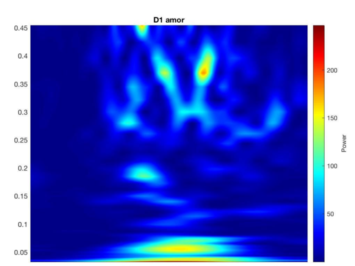
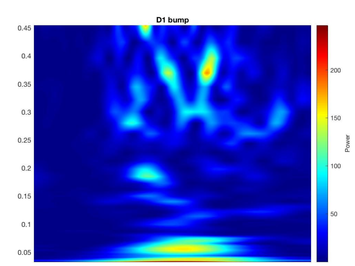
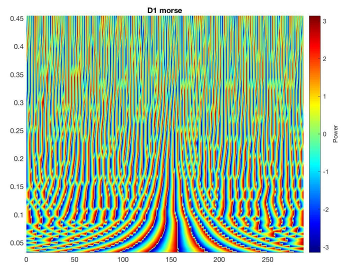
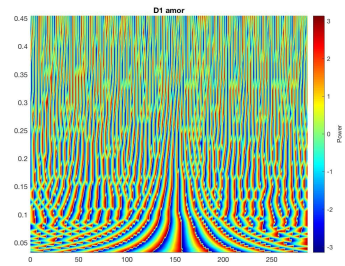
		
Morse Click D1	Gabor Click D1	Bump Click D1
		
Morse Conversion D1	Gabor Conversion D1	Bump Conversion D1

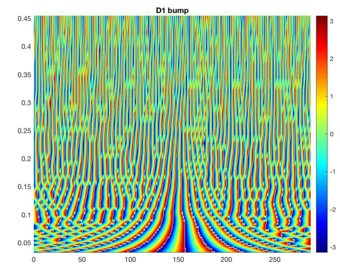
Fig 6. Morse, Gabor, and bump wavelet magnitude with abnormality in D1 where the time unit is 5min



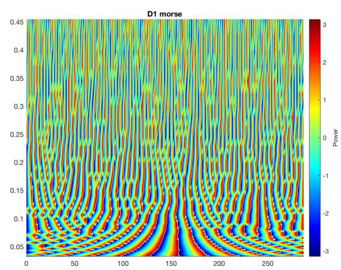
Morse Click D1



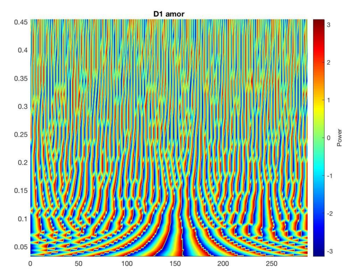
Gabor Click D1



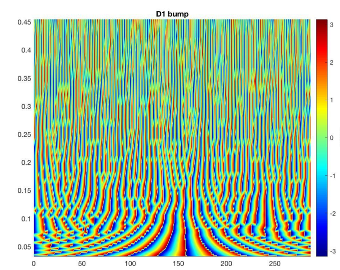
Bump Click D1



Morse Conversion D1



Gabor Conversion D1



Bump Conversion D1

Fig 7. Morse, Gabor, and bump wavelet magnitude with abnormality in D1 where the time unit is 5min

Experiments

In the first three week click data, it is shown that Day 1, 11, 12, 18, and 20 have abnormal pattern and other days do not. A set of Day 1, 11, 12, 18 and another set of Day 2 to 10 are configured as abnormality train data and normal train data respectively. After the training AlexNet with the small amount of data, performance test is proceeded. The test accuracy is **0.8571428571428571** at most and 0.5714285714285714 at least. All the data used for both train and test can be downloaded here.

Wavelet	**Data type**	**Time unit**	**Test accuracy**
Morse	Magnitude	1min	0.8571428571428571
Gabor	Magnitude	1min	0.8571428571428571
Bump	Magnitude	1min	0.8571428571428571
Morse	Magnitude	5min	0.6666666666666666
Gabor	Magnitude	5min	0.6666666666666666
Bump	Magnitude	5min	0.5714285714285714

Table 1. Results

AlexNet

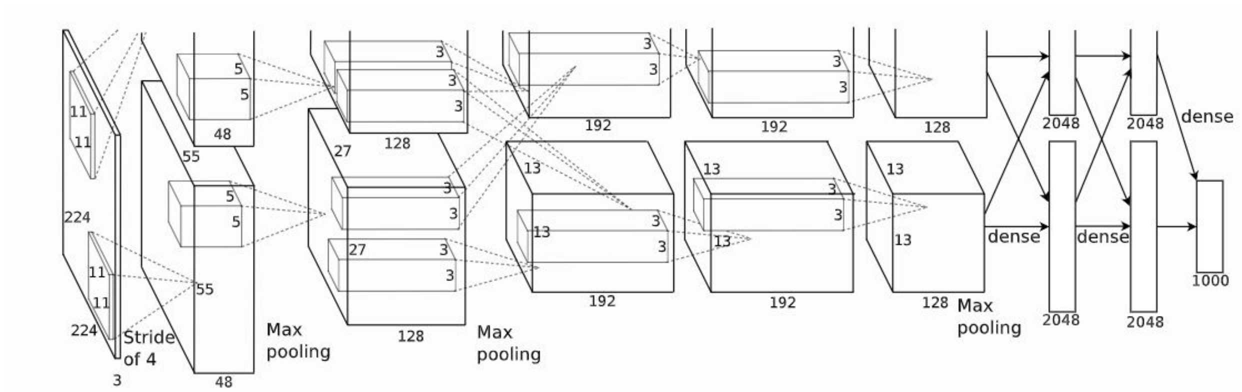
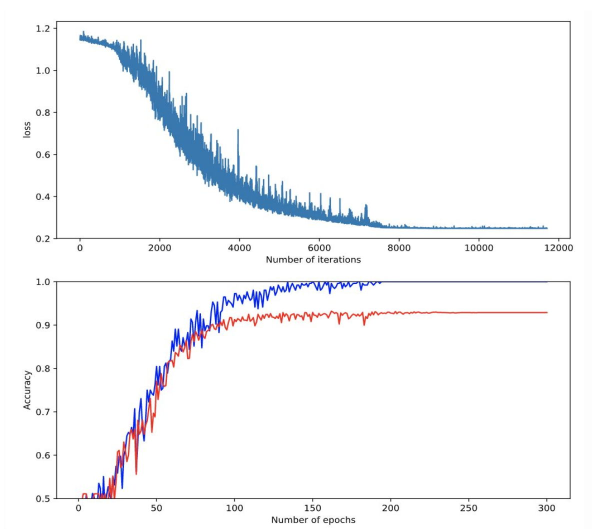


Fig 8. Architecture of AlexNet

In the AlexNet paper, the model was proposed to have a local response normalization (LRN) layer just after some convolutional and pooling layers, so that the output at each layer is constantly adjusted. Today, because of improvements in activation functions and the emergence of batch normalization methods, LRN is rarely used in modern convolutional neural network architectures and is not good at simplifying implementation. Thus, in this work, we dropped LRN.



All the work was coded in python which can be downloaded here.

Fig 9. AlexNet learning rate and loss/ learning rate and accuracy

4. Conclusion

Regarding the angle phase data, most of the images do not contain any unique features that stands out for a CNN to detect quick and easy. The pattern of vertical stripes are definitely noticeable in human eye though.

Also, due to the lack of training datasets AlexNet could not detect abnormality high.

However, in anomaly detection with combination of wavelet and AlexNet, it is found out that 1) signal analysis with less time unit (or low frequency) works more powerful than larger time unit (or higher

frequency). Also, 2) if we figure ways out to denoise convolution data then it will be able to enhance a CNN’s performance.
0.857 accuracy is not low when considering the size of training data. Thus, with more training data and a denoising method in wavelet transform, combination of wavelet and AlexNet will perform better.

Also, except for the angles and magnitudes, we still have other options. Wavelet coefficients or X-axis and Y-axis values (or 3 dimensions of time, X-axis, and Y-axis values) can be used to let the model learn and detect abnormality.