# Deep Learning and Spiking Neural Networks for Advanced Data Mining

**Project Report**

Emanuele Artioli
12010733

Alpen-Adria Universität Klagenfurt

*June 29, 2021*

# Contents

# 1 Introduction

The aim of this project is the interpretation of domestic sensor patterns to predict the type of activity performed. Leveraging the google news pre-trained model for word embedding, the model developed should not only be capable of returning activities it was trained on, but to infer the meaning of new patterns and translate them into activities never seen.

# 2 Data Description

Features.csv is a file of sensor events that represent home activities. Every feature represents a sensor, and every entry contains the name of the activity with the amount of times every sensor was activated during that activity. An example below:

| | Activity | D001 | D002 | D003 | M001 | M002 | M003 | M004 | M005 | M006 | ... | M015 | M016 | M017 | MA018 | MA019 | MA020 | MA021 | MA022 | MA023 | MA024 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 915 | Enter Home | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1316 | Groom | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 |
| 1949 | Sleep | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 |
| 2598 | Wash Dinner Dishes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| 1189 | Groom | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 |
| 2410 | Toilet | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 |
| 565 | Dress | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 0 | 0 |
| 2847 | Watch TV | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1911 | Phone | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 8 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1856 | Personal Hygiene | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |

Figure 1: sample of Features.csv

The shape of Features.csv is: 3231 rows $\times$ 28 columns.

# 3 Data Selection and Preparation

A preliminary analysis on the data shows a number of empty columns, sensors that were never activated during any kind of activity and therefore don't contribute to any information. These columns were dropped to reduce computation. After dropping these columns, the shape is: 3231 rows $\times$ 17 columns.
Apart from this, the data consists of sparse integers spanning from 0 to 597. This vast range is problematic since it can cause smaller values to lose significance, whereas domain knowledge would suggest a sensor activated just a couple times is already a significant data point. it is therefore optimal to normalize the data.

# 4 Problem Analysis

In the development of a successful model, the following challenges have to be overcome:

### 4.1 Train Test Split

The classic split in train and test data has to be revised since the labels present in the test data should not appear in the train data, i.e. the test data should be composed of new classes.

### 4.2 Embedding

To be able to predict unseen labels, the model will train not on the activities, but on the 300-vector embedding of them, therefore instead of a classification problem we are faced with a multivariate regression problem, much more complex.

### 4.3 Reverse Embedding

The predictions of the embedded words are not gonna be exact, therefore it's necessary to find the embedding of the word in our vocabulary that most closely resembles the prediction, and consider that word as the actual prediction of the model. Comparing the predictions so derived with the test values will allow the evaluation of the model with the classic metrics of accuracy, precision, recall and F-measure.

## 5 Mathematical Description of the Proposed Models

The first densely connected model has form:

$$
\begin{gathered}
A^{(1)} = relu(X^T W^{(1)} + \bar{b}^{(1)}), W[16, 300] \\
A^{(2)} = relu(A^{(1)T} W^{(2)} + \bar{b}^{(2)}), W[300, 300] \\
L = \frac{1}{N} \sum_{i=1}^{N} ||\bar{y} - A^{(2)}||^2 + ||W^{(2)}||^2
\end{gathered}
$$

This simple model is not capable of modeling the complexity of 300 labels, after many incremental changes the final model has become:

$$
\begin{gathered}
A^{(1)} = elu(X^T W^{(1)} + \bar{b}^{(1)}), W[16, 16] \\
A^{(2)} = \frac{X-\mu}{\sqrt{\sigma^2}} \\
A^{(3)} = elu(A^{(2)T} W^{(3)} + \bar{b}^{(3)}), W[16, 32] \\
A^{(4)} = \frac{X-\mu}{\sqrt{\sigma^2}} \\
A^{(5)} = elu(A^{(4)T} W^{(5)} + \bar{b}^{(5)}), W[32, 64] \\
A^{(6)} = \frac{X-\mu}{\sqrt{\sigma^2}} \\
A^{(7)} = elu(A^{(6)T} W^{(7)} + \bar{b}^{(7)}), W[64, 128] \\
A^{(8)} = \frac{X-\mu}{\sqrt{\sigma^2}} \\
A^{(9)} = elu(A^{(8)T} W^{(9)} + \bar{b}^{(9)}), W[128, 256] \\
A^{(10)} = \frac{X-\mu}{\sqrt{\sigma^2}} \\
A^{(11)} = elu(A^{(10)T} W^{(11)} + \bar{b}^{(11)}), W[256, 512] \\
A^{(12)} = \frac{X-\mu}{\sqrt{\sigma^2}}
\end{gathered}
$$

$$A^{(13)} = elu(A^{(12)T}W^{(13)} + \bar{b}^{(13)}), W[512, 1024]$$
$$A^{(14)} = \frac{X-\mu}{\sqrt{\sigma^2}}$$
$$A^{(15)} = elu(A^{(14)T}W^{(15)} + \bar{b}^{(15)}), W[1024, 300]$$
$$L = \frac{1}{N}\sum_{i=1}^{N} ||\bar{y} - A^{(15)}||^2 + ||W^{(16)}||^2$$

# 6 Performance and Model Evaluation

The starting performance was underwhelming. 36% of accuracy doesn't seem terrible until we check the confusion matrix to realize that the whole test set has been identified in the same class.

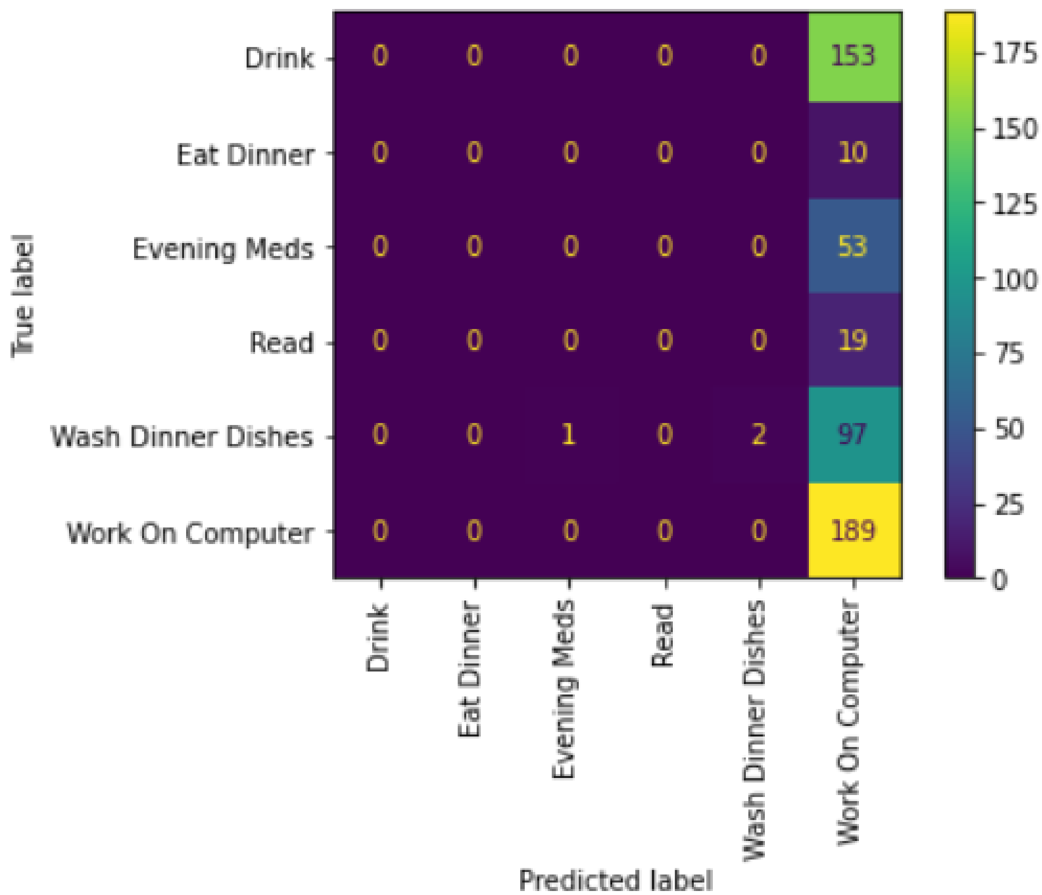


Figure 2: confusion matrix of baseline model

After a great number of modifications, oversampling the classes showed good signs of improvements, and after coupling this with the binarization of the data (setting to 1 every non null value, so that the information from the data is no more how many times each sensor was activated in the activity, but only if it were activated or not) the model

offers an accuracy higher than 90%. Unfortunately, this solution didn't prove to be very consistent.
The focus was then moved to the words chosen for the test set: words that have in common a good representation and have also highly represented similar words in the training set.
After some trial and error tweaking the hyperparameters, the final configuration was identified in using the ELU activation function on the previous model. This model achieves consistent predictions: on the right set of test words, the accuracy was shy of 98%.

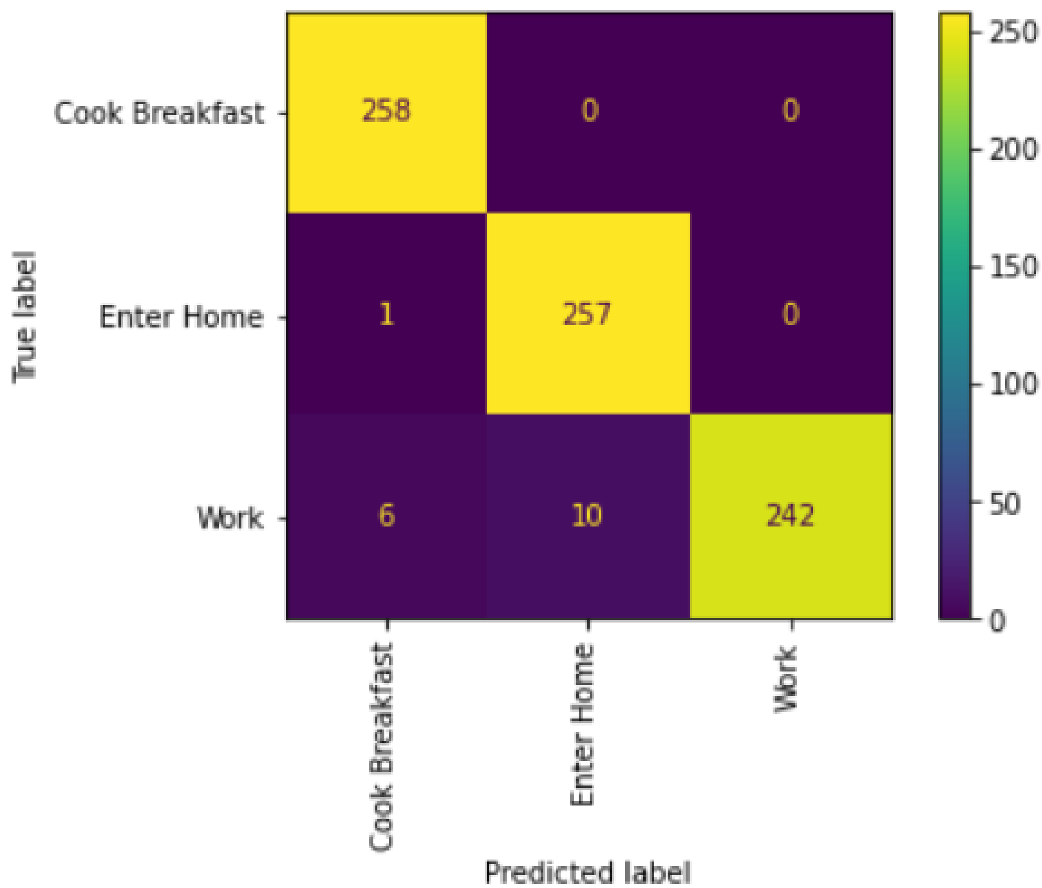


Figure 3: confusion matrix of final model

# 7 Results and Interpretation

The final model achieves particularly high levels of accuracy. In particular, the set of test words ['Cook Breakfast', 'Enter Home', 'Work'] achieves 98% accuracy. Adding words to the test set reduces the accuracy consistently, but it has to be considered the difficulty in finding the right test set: if a test word has not a high and close representation in the training set, it's clearly very difficult for the model to learn its distribution. Words with a good enough representation are very precisely predicted. Interestingly though, the model is not bothered by the presence of noise words, those that aren't correctly predicted: it is still able to pick the represented ones from the others and from each other, and classifies the other ones with great confidence and variety (e.g. The activity 'Drink' is confidently predicted as 'Wash Dinner Dishes'. It is not spread over many classes. Also, no other activity is predicted as 'Wash Dinner Dishes'). This is not trivial, as it means that not only the network picks up on the distributions of data, but also that the different words have a significantly different pattern that the network can identify even when it is not able to correctly classify them.

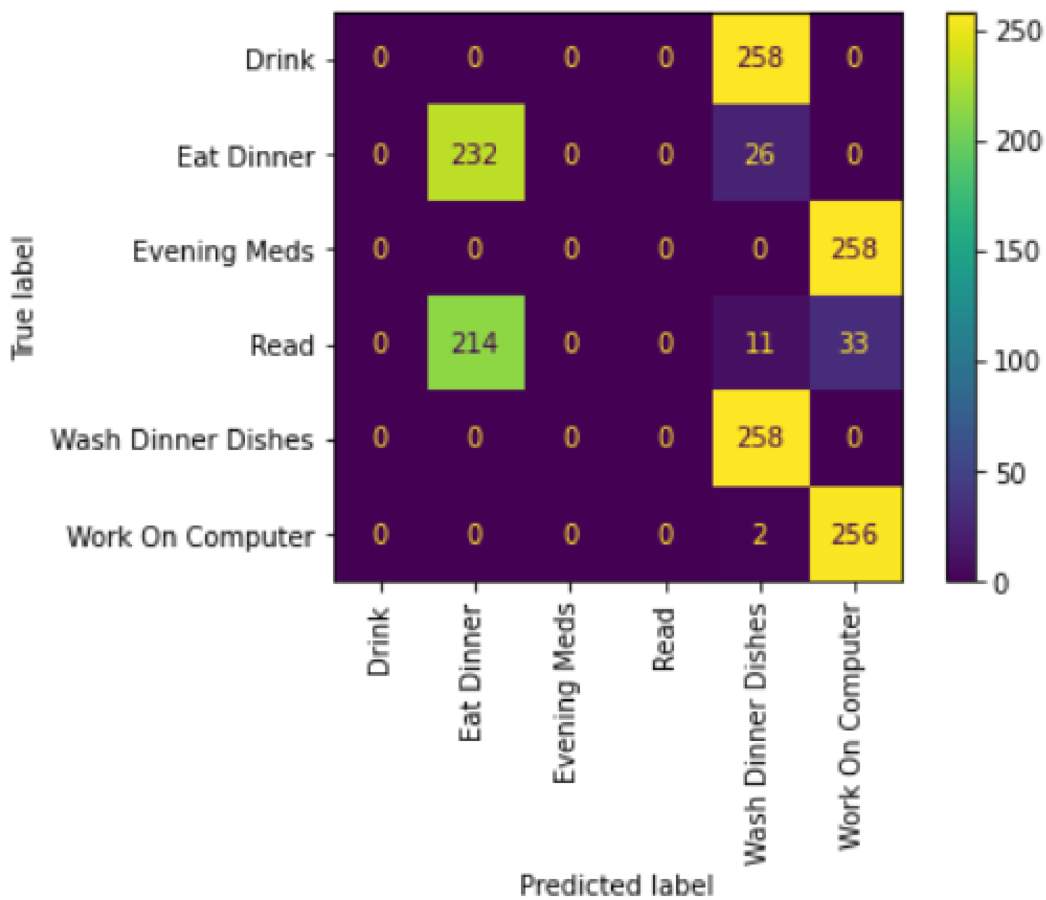


Figure 4: confusion matrix of final model with random test set

# 8 Conclusions

This model can be considered successful. The performance could be improved greatly by recording more data and tracking a higher variety of activities so that the classes are numerous enough to not require oversampling, which introduces a great deal of noise and every class has a sufficient number of similar words.