|  |  |  |
| --- | --- | --- |
| **CISC/CMPE 452/COGS 400/CISC 874 Neural and Genetic Cognitive Models** | | |
| Course Project: Prediction of Glucose Concentration in Type 1 Diabetes Patients |

Evan Cloutier, 10091086

Ryan Dick, 10092399 (Group Rep)

Aidan Gunda, 10060756

Jordan Nanos, 10100261

Nolan Nisbet, 10089873

# Solution Overview

## 1.1 Approach

## 1.2 Justification

One page description of the solution approach justifying why you took this approach and what tool was used for analytics.

* Explain why we used overnight data
* Explain Feed Forward network architecture
  + 2 hidden layer
  + 15 nodes in each layer
  + linear output functions(rather than a bounded function like a sigmoid or tanh)
  + trained using per-epoch Adam Optimizer
    - mention that Adam is an extension of gradient descent (I think), but it produced better results than basic gradient descent
    - Cost function was mean squared error
* Explain Recurrent neural network
  + Used a single LSTM cell
  + LSTM cell output was multiplied by a weight and a bias was added to get the output
  + RNN made decisions based on last 7 inputs
  + Same as for FFN:
    - linear output functions used
    - Adam optimizer
    - MSE cost function

## 1.3 Data Processing

Also submit a flow diagram to show the steps of data processing (cleansing, normalizing, encoding, processing, etc.)

* Data Extraction:
  + Group data based on patient ID
  + Organize points into 24 hour windows (noon to noon)
  + Save each day worth of data to a separate file within a folder named with the patient ID
* Data preprocessing:
  + For each 24-hour data file in a patient folder:
    - Sort the data by time (just in case its unsorted)
    - Extract only the overnight data
    - Find windows containing 7 consecutive measurements at 10 minute intervals followed by a measurement at a 20 minute predicition horizon.
      * Data was not always collected at regular intervals, so perform a linear interpolation between points if necessary, but never interpolate between points that are more than 12 minutes apart.
    - If a valid window is found, write a single data point to the output file containing the seven 'input' measurements and the single desired output

# 2.0 Results

Present the results and based on your validation criteria, show how well your system is doing. For work on classification you should include classification accuracy as shown by the MATLAB confusion matrix, precision and recall measures.

## 2.1 Feed Forward Neural Network Results

Table : Evaluation criteria for FFNN trained on Patient 149 Data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Patient** | **MSE** | **% of Lows Identified** | **% of Highs Identified** | **Number of False Lows** | **Number of False Highs** |
| ***149*** | *249* | *79%* | *93%* | *14* | *44* |
| **151** | 100 | 75% | 96% | 8 | 5 |
| **174** | 133 | 77% | 88% | 4 | 11 |

Table : Evaluation criteria for FFNN trained on Patient 151 Data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Patient** | **MSE** | **% of Lows Identified** | **% of Highs Identified** | **Number of False Lows** | **Number of False Highs** |
| **149** | 278 | 72% | 93% | 19 | 48 |
| ***151*** | *101* | *82%* | *96%* | *11* | *4* |
| **174** | 154 | 69% | 86% | 5 | 17 |

Table : Evaluation criteria for FFNN trained on Patient 174 Data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Patient** | **MSE** | **% of Lows Identified** | **% of Highs Identified** | **Number of False Lows** | **Number of False Highs** |
| **149** | 270 | 84% | 93% | 19 | 49 |
| **151** | 105 | 73% | 97% | 10 | 3 |
| ***174*** | *143* | *77%* | *87%* | *4* | *14* |

## 2.2 Recurrent Neural Network Results

Table : Evaluation criteria for RNN trained on Patient 149 Data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Patient** | **MSE** | **% of Lows Identified** | **% of Highs Identified** | **Number of False Lows** | **Number of False Highs** |
| **149** | 279 | 75% | 93% | 14 | 46 |
| **151** | 106 | 80% | 97% | 13 | 7 |
| **174** | 161 | 69% | 87% | 5 | 16 |

Table : Evaluation criteria for RNN trained on Patient 151 Data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Patient** | **MSE** | **% of Lows Identified** | **% of Highs Identified** | **Number of False Lows** | **Number of False Highs** |
| **149** | 314 | 67% | 92% | 12 | 54 |
| **151** | 105 | 70% | 96% | 9 | 5 |
| **174** | 173 | 69% | 85% | 5 | 17 |

Table : Evaluation criteria for RNN trained on Patient 174 Data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Patient** | **MSE** | **% of Lows Identified** | **% of Highs Identified** | **Number of False Lows** | **Number of False Highs** |
| **149** | 408 | 70% | 93% | 16 | 51 |
| **151** | 112 | 65% | 97% | 12 | 9 |
| **174** | 151 | 77% | 89% | 3 | 15 |

## Discussion:

* MSE really isn't low enough. Would like to get down to like 30
  + This probably won't be possible without more information like food, insuling, activity etc.
* It didn't matter that much which patient the network was trained on. Unfortunately, this suggests that a static algorithm may be equally well-suited to the problem as a neural network approach.
* The amount of data was not sufficient to achieve excellent high and low detection:
  + Despite taking ~30 days worth of data, there was only a handful of lows and highs present for training and testing. (Most of the data was in the 'normal' range.
* Number of false lows and false highs were higher than desired:
  + This could be addressed in the future by developing a cost function that penalizes false lows/highs more strongly. Such a cost function is very difficult to implement using tensorflow's API.
* The FFNN and RNN gave very similar results. It was observed that the FFNN provided slightly more consistent training. In other words, trainging the FFNN multiple times consistently produced a very similar MSE, whereas training the RNN multiple times resulted in fluctuating MSE values.

# 3.0 Contributions of Each Team Member

|  |  |
| --- | --- |
| **Group Member** | **Contribution** |
| Evan Cloutier, 10091086 |  |
| Ryan Dick, 10092399 (Group Rep) | Data Cleansing and Processing |
| Aidan Gunda, 10060756 |  |
| Jordan Nanos, 10100261 |  |
| Nolan Nisbet, 10089873 | Data Cleansing and Processing |

# 4.0 Project Code

Submit the code with paragraph style and in-line comments about what each part of the code is doing. May be you can later extend this work for one of your project courses.