Why MultiLayer Perceptron/Neural Network?

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.

Other advantages include:

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
2. One of the preferred techniques for gesture recognition.
3. MLP/Neural networks do not make any assumption regarding the underlying probability density functions or other probabilistic information about the pattern classes under consideration in comparison to other probability based models [1].
4. They yield the required decision function directly via training.
5. A two layer backpropagation network with sufficient hidden nodes has been proven to be a universal approximator [2] [3].

Objective:

The Neural network was implemented to recognize the three hand gestures namely grasp, point and push, irrespective of who is doing these hand gestures.

Attributes:

Attribute Extraction:

- The first step was to extract relevant attributes from the data. Following attributes were extracted for neural nets:
  1. Sumdis = Sum of the distances of 3 finger markers from the centroid, figure 2.
  2. SumdeltaX = Sum of the differences of X displacement of each finger marker with respect to the X displacement of centroid, figure 1.
  3. SumdeltaY = Sum of the differences of Y displacement of each finger marker with respect to the Y displacement of centroid, figure 1.
  4. SumdeltaZ = Sum of the differences of Z displacement of each finger marker with respect to the Z displacement of centroid, figure 1.
Next step was attribute selection. Initially Principal Component Analysis was used for attribute selection, but a poor discrimination between gestures was observed when the data was projected on the PCA axes. Below are the results from PCA and the figure depicting the projections on PCA axes (figures were generated using WEKA, which does not have feature to save figures, hence screen shots are provided):

Instances: 3821
Attributes: 5
   Attribute Selection on all input data

Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>-0.07</th>
<th>0.57</th>
<th>0.22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.07</td>
<td>1</td>
<td>0.54</td>
<td>-0.71</td>
</tr>
<tr>
<td>-0.07</td>
<td>1</td>
<td>0.54</td>
<td>1</td>
<td>-0.17</td>
</tr>
<tr>
<td>0.57</td>
<td>0.54</td>
<td>1</td>
<td>-0.17</td>
<td>1</td>
</tr>
<tr>
<td>0.22</td>
<td>-0.71</td>
<td>-0.17</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

eigenvalue proportion cumulative
eigenvalue proportion cumulative

-0.654SumdeltaY+0.529Sumdis-0.525SumdeltaZ-0.131SumdeltaX
-0.744SumdeltaX-0.479SumdeltaZ-0.426Sumdis+0.189SumdeltaY
-0.651Sumdis+0.559SumdeltaX-0.411SumdeltaZ-0.309SumdeltaY

Eigenvectors

V1 V2 V3

-0.1309 -0.744 0.5586 SumdeltaX
-0.6543 0.189 -0.3087 SumdeltaY
-0.5246 -0.4791 -0.4107 SumdeltaZ
0.5287 -0.4257 -0.6512 Sumdis

Ranked attributes:

0.5023 1 -0.654SumdeltaY+0.529Sumdis-0.525SumdeltaZ-0.131SumdeltaX
0.1244 2 -0.744SumdeltaX+0.479SumdeltaZ-0.426Sumdis+0.189SumdeltaY
0.0336 3 -0.5246 -0.4791 -0.4107 SumdeltaZ
0.651 Sumdis+0.559SumdeltaX-0.411SumdeltaZ-0.309SumdeltaY
X axis   -0.651Sumdis+0.559SumdeltaX-0.411SumdeltaZ-0.309SumdeltaY
Y axis   -0.654SumdeltaY+0.529Sumdis-0.525SumdeltaZ-0.131SumdeltaX

Therefore to decide upon the attributes heuristics, observations and domain knowledge were used and three attributes namely Sumdis, SumdeltaY and SumdeltaZ were selected.

Models:

From the selected features two models were created one with 2 dimensional feature space (Sumdis and SumdeltaY) and a 3 dimensional feature space (Sumdis, SumdeltaY and SumdeltaZ). Two models were created to observe the effect of dimensionality (increasing the number of features) and also after the classification results from 2D features space model it was observed that increasing a feature i.e. SumdeltaZ (which was logical as there is a major difference between the displacements along Z, of Grasp and Point gestures). All the inputs to the neural net were normalized in the range of -1 to 1.

Model paramenters:-

- Parameters for MLP using 2D feature space (figure 3)
  - No. of Hidden layer 1 with 2 hidden nodes,
  - Learning rate = 0.3, Momentum = 0.2,
  - Epochs =600, sigmoid for activation.
Attributes:  2
Sumdis
SumdeltaY

### Classifier model (full training set) ###

**Sigmoid Node 0**

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node 0</td>
<td>-2.873084657868268</td>
</tr>
<tr>
<td>Node 3</td>
<td>11.098430581589703</td>
</tr>
<tr>
<td>Node 4</td>
<td>1.6427091491248684</td>
</tr>
</tbody>
</table>

**Sigmoid Node 1**

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node 0</td>
<td>-20.95588207113826</td>
</tr>
<tr>
<td>Node 3</td>
<td>-15.325529796449096</td>
</tr>
<tr>
<td>Node 4</td>
<td>22.676419665124076</td>
</tr>
</tbody>
</table>

**Sigmoid Node 2**

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node 0</td>
<td>4.462354683446656</td>
</tr>
<tr>
<td>Node 3</td>
<td>-6.567587151859792</td>
</tr>
<tr>
<td>Node 4</td>
<td>-7.1281632679527664</td>
</tr>
</tbody>
</table>

**Sigmoid Node 3**

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node 0</td>
<td>71.06468600599895</td>
</tr>
<tr>
<td>Node 3</td>
<td>-6.567587151859792</td>
</tr>
<tr>
<td>Node 4</td>
<td>-7.1281632679527664</td>
</tr>
</tbody>
</table>

**Sigmoid Node 4**

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node 0</td>
<td>53.054453951951984</td>
</tr>
<tr>
<td>Node 3</td>
<td>-6.567587151859792</td>
</tr>
<tr>
<td>Node 4</td>
<td>-7.1281632679527664</td>
</tr>
</tbody>
</table>

- **Class Gesture-Grasp**
  - Input
  - Node 0

- **Class Gesture-Point**
  - Input
  - Node 1

- **Class Gesture-Push**
  - Input
  - Node 2

- **Parameters for MLP using 3D feature space (figure 4)**
  - No. of Hidden layer 1 with 3 hidden nodes,
  - Learning rate = .12, Momentum = 0.2,
  - Epochs = 600, sigmoid for activation.

Attributes:  3
Sumdis
SumdeltaY
SumdeltaZ

### Classifier model (full training set) ###

**Sigmoid Node 0**

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node 0</td>
<td>5.83055323849687</td>
</tr>
<tr>
<td>Node 3</td>
<td>23.316296519360424</td>
</tr>
<tr>
<td>Node 4</td>
<td>-6.451972450033389</td>
</tr>
<tr>
<td>Node 5</td>
<td>-25.68306054604823</td>
</tr>
</tbody>
</table>

**Sigmoid Node 1**

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node 0</td>
<td>-3.854105033173585</td>
</tr>
<tr>
<td>Node 3</td>
<td>-16.492886626269794</td>
</tr>
<tr>
<td>Node 4</td>
<td>-15.001763796835775</td>
</tr>
</tbody>
</table>
The above mentioned parameters were derived after experimenting with various parameters, using these the best classification was achieved.

**Results:**

- 2D Feature space
  - 10 folds cross validation on Manu’s Data
Correctly Classified Instances 3171 80.1567 %
Incorrectly Classified Instances 785 19.8433 %
Root mean squared error 0.3058
Kappa statistic 0.7023
Mean absolute error 0.1929
Relative absolute error 43.4131 %
Root relative squared error 64.8791 %
Total Number of Instances 3956

== Detailed Accuracy By Class ==

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gesture-Grasp</td>
<td>0.689</td>
<td>0.076</td>
<td>0.82</td>
<td>0.689</td>
<td>0.749</td>
</tr>
<tr>
<td>Gesture-Point</td>
<td>0.804</td>
<td>0.161</td>
<td>0.718</td>
<td>0.804</td>
<td>0.759</td>
</tr>
<tr>
<td>Gesture-Push</td>
<td>0.914</td>
<td>0.062</td>
<td>0.879</td>
<td>0.914</td>
<td>0.896</td>
</tr>
</tbody>
</table>

Model trained on Manu’s data and tested on Rita’s data

Correctly Classified Instances 2622 71.9144 %
Incorrectly Classified Instances 1024 28.0856 %
Kappa statistic 0.5825
Mean absolute error 0.2159
Root mean squared error 0.3628
Relative absolute error 48.5867 %
Root relative squared error 76.9631 %
Total Number of Instances 3646

== Detailed Accuracy By Class ==

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gesture-Grasp</td>
<td>0.447</td>
<td>0.09</td>
<td>0.748</td>
<td>0.447</td>
<td>0.559</td>
</tr>
<tr>
<td>Gesture-Point</td>
<td>0.792</td>
<td>0.307</td>
<td>0.547</td>
<td>0.792</td>
<td>0.647</td>
</tr>
<tr>
<td>Gesture-Push</td>
<td>0.977</td>
<td>0.023</td>
<td>0.95</td>
<td>0.977</td>
<td>0.963</td>
</tr>
</tbody>
</table>

2D feature space - Manu (10Folds X validation)  2D feature space test on Rita’s data
Confusion Matrix
\begin{tabular}{ccc} 
\text{a} & \text{b} & \text{c} \tableline 
910 & 336 & 75 \text{a = Gesture-Grasp} \\
173 & 1073 & 89 \text{b = Gesture-Point} \\
27 & 85 & 1188 \text{c = Gesture-Push} \\
\end{tabular} \hspace{1cm} \\
\begin{tabular}{ccc} 
\text{a} & \text{b} & \text{c} \tableline 
610 & 741 & 15 \text{a = Gesture-Grasp} \\
200 & 919 & 42 \text{b = Gesture-Point} \\
5 & 21 & 1093 \text{c = Gesture-Push} \\
\end{tabular}

- 3D Feature space

- \textbf{10 folds cross validation on Manu’s Data}
  - Correctly Classified Instances: 3511, 91.8869 %
  - Incorrectly Classified Instances: 310, 8.1131 %
  - Kappa statistic: 0.8779
  - Mean absolute error: 0.0916
  - Root mean squared error: 0.2112
  - Relative absolute error: 20.6521 %
  - Root relative squared error: 44.8404 %
  - Total Number of Instances: 3821

  \begin{tabular}{cccccc} 
  TP Rate & FP Rate & Precision & Recall & F-Measure & Class \tableline 
  0.945 & 0.043 & 0.92 & 0.945 & 0.932 & Gesture-Grasp \\
  0.92 & 0.055 & 0.9 & 0.92 & 0.91 & Gesture-Point \\
  0.888 & 0.024 & 0.941 & 0.888 & 0.914 & Gesture-Push \\
  \end{tabular}

- \textbf{Model trained on Manu’s data and tested on Rita’s data}
  - Correctly Classified Instances: 2463, 68.953 %
  - Incorrectly Classified Instances: 1109, 31.047 %
  - Kappa statistic: 0.5349
  - Mean absolute error: 0.2277
  - Root mean squared error: 0.4343
  - Relative absolute error: 51.3569 %
  - Root relative squared error: 92.2682 %
  - Total Number of Instances: 3572

  \begin{tabular}{cccccc} 
  TP Rate & FP Rate & Precision & Recall & F-Measure & Class \tableline 
  0.493 & 0.15 & 0.67 & 0.493 & 0.568 & Gesture-Grasp \\
  0.714 & 0.322 & 0.516 & 0.714 & 0.599 & Gesture-Point \\
  0.919 & 0 & 1 & 0.919 & 0.958 & Gesture-Push \\
  \end{tabular}
3D feature space - Manu (10Folds X validation)  

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>&lt;-- classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>1248</td>
<td>62</td>
<td>11</td>
<td>Gesture-Grasp</td>
</tr>
<tr>
<td>53</td>
<td>1228</td>
<td>54</td>
<td>Gesture-Point</td>
</tr>
<tr>
<td>55</td>
<td>75</td>
<td>1035</td>
<td>Gesture-Push</td>
</tr>
</tbody>
</table>

3D feature space test on Rita’s data

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>&lt;-- classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>674</td>
<td>692</td>
<td>0</td>
<td>Gesture-Grasp</td>
</tr>
<tr>
<td>332</td>
<td>829</td>
<td>0</td>
<td>Gesture-Point</td>
</tr>
<tr>
<td>0</td>
<td>85</td>
<td>960</td>
<td>Gesture-Push</td>
</tr>
</tbody>
</table>

Feature space plots:

X axis - Sumdis Vs Y axis - SumdeltaY
Grasp Point Push
Conclusions: -

- Adding extra feature i.e. increasing dimensionality does not help in this case.
- In comparison to 2 features, though good results were observed for 10 folds X validation, but the performance degraded for test data (look like overfitting).
- For 3 features more point gestures were misclassified as grasp, but more grasp gestures were misclassified as point for 2 features. A tradeoff between increasing the total classification accuracy and true positives can be seen, the neural net was not able to optimize this.
- In both cases gesture Push was unambiguously recognized with True positive rate as high as 0.977.
For MLP deciding upon learning rate is very important, a lower rate performed better in 3D feature spaces.

After experimenting with different number of hidden layer and hidden node, it was found that a single hidden layer with few hidden nodes performed better. Adding extra hidden layer does not help always, but increasing the number of nodes might help.

**Discussion & future work:**

Their ability to learn by example makes neural nets very flexible and powerful. There is no need to devise an algorithm in order to perform a specific task; i.e. there is no need to understand the internal mechanisms of that task. Along various other advantages of Neural nets there disadvantages too they cannot be programmed to perform a specific task; the examples must be selected carefully otherwise useful time is wasted or even worse the network might be functioning incorrectly. Also, network finds out how to solve the problem by itself, hence its operation can be unpredictable. The problem with the backpropagation algorithm is that it tries to find a local minimum in the error function output, if it ends up in finding the wrong one, the results can drastically bad, and that’s why learning rate is important. Instead of using a simple backpropagation algorithm advanced algorithms like hyper rectangular composite NN (HRCNN) using supervised decision directed learning (SDDL) can be used [1]. More appropriate features can be extracted. And a well modeled neural net can be developed for real time hand gesture recognition.

**Reference:**