Preference Neural Network and its Applications

by **Ayman Ahmed Elgharabawy**

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Dr. Mukesh Prasad

University of Technology Sydney
School of Computer Science
Faculty of Engineering and Information Technology

December 2021

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Ayman Ahmed ELgharabawy* declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Computer Science*, *Faculty of Engineering and Information Technology.* at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

SIGNATURE

Production Note:
Signature removed prior to publication.

DATE: 6TH DECEMBER, 2021

ABSTRACT

This thesis proposes a novel label ranker network to learn the relationship between labels for classification and ranking problems. The Preference Neural Network (*PNN*) uses *spearman* correlation gradient ascent and two new activation functions,namely positive smooth staircase (*PSS*) and smooth staircase (*SS*) that accelerate the ranking by creating deterministic preference values. *PNN* is proposed in two forms, fully connected simple layers and Preference Net (*PN*), where the latter is the deep ranking form of *PNN* to learning feature selection using a novel ranker kernel to solve images classification problem. *PN* uses a new type of multiple size weighted ranker kernel to generate a feature map. *PNN* outperforms five previously proposed methods for label ranking, obtaining state-of-the-art results on label ranking, and *PN* achieves promising results on *CFAR-10* with high computational efficiency. The thesis includes different types of *PN* architecture to solve the problem of subgroup label ranking. Subgroup label ranking, which aims to rank labels in individual groups using a single ranking model, is a new problem in preference learning. This thesis also introduces the subgroup preference neural Network (*SGPNN*) that combines multiple networks that have different activation functions, learning rate, and output layer into one to discover the hidden relation between the subgroups' multi-labels. The *SGPNN* is a feedforward (*FF*), partially connected that has a single middle layer and uses stairstep (*SS*) multi-valued activation function to optimize learning and achieve better ranking performance. The novel structure of the proposed *SGPN* consists of a multi-activation function neuron (*MAFN*) in the middle layer to rank each subgroup independently. The *SGPNN* uses gradient ascent to maximize the Spearman rank correlation between the subgroups' multi-labels. Each label is represented by an output neuron that has a single *SS* function. Experiments were conducted that applied the *SGPNN* to a new synthesized dataset with subgroup label ranking achieving promising results in computational cost and performance. *PN* has experimented on image recognition benchmarks datasets, and *SG-*

PNN is applied on *EEG* motor imagery BCI competition IV dataset 2b to solve the data ambiguity. Under these varying backgrounds and scenarios, the thesis has shown that the proposed *PNN* provides a learning tool for label ranking and class classification. *PNN* outperforms label ranking state-of-the-art and gives promising results in image classification via literature explanation and empirical results.

DEDICATION

To my mothers' soul who died during Covid-19 pandemic.

To my wife and my kids who supported me and suffered to live without me during this research.

To my supervisors Dr.Mukesh Prasad and Prof. Ct. Lin who believed in my ability to finish this research after all these hardships.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my supervisors *Dr. Mukesh Prasad* and *Prof. Chin Teng Lin* for the continues support of my Ph.D study and research, for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis.

I also would like to appreciate my fellow in computer science: *Dr. Mohamed Barbar* for his code review and performance optimization. and my fellow *Dr. Akram Youssry* in center of quantum for his fruitful discussion in the design of the activation function.

I also would like to thank *Prof. Nikhil Pal* in Indian Statistical Institute for his valuable suggestions in choosing the error function of Preference Neural Network to make it a unique and novel peace of network to reach the state-of-the-art in the machine learning.

additional to, I like to thank the *Dr. Nabin Sharma* in department of drones for his valuable comments in image recognition to validate the results of preference net.

Last but not the least, I would like to thank my family: my wife and my kids, for their unconditional support, both financially and emotionally throughout the whole PhD journey.

Ayman Elgharabawy

December 2021 @UTS

PUBLICATIONS

JOURNAL PUBLICATIONS RELATED TO THE THESIS :

1) Preference Neural Network. **(Chapter Two)** (Under Review) (Elgharabawy et al. 2021*a*)

2) Preference Net: Image Recognition using Ranking Reduction to Classification. **(Chapter Three)** (Under Review)

3) Subgroup Preference Neural Network. **(Chapter Four)** (Published) (Elgharabawy et al. 2021*b*)

Table of Contents

List of Figures **xiii**

List of Tables **xviii**

1 Introduction **1**

1.1 Problem Statement 1

1.2 Research Scope 2

1.3 Research Questions 2

1.4 Research Methods 3

1.5 Aims and Objectives 4

1.6 Research Challenges 5

1.6.1 Combining the *PL* domain and *ANN* 5

1.6.2 Combining the relation between *LP* domain and *DL* domain 6

1.6.3 Combining the relation between *SD* domain and *LR* . 6

1.7 Research Contributions 6

1.8 Thesis Structure 8

2 Preference Learning and Neural Network **11**

2.1 Introduction 11

2.2 Label Ranking (*LR*) 12

2.2.1 Restricted Label Ranking 13

2.2.2 Unrestricted Label Ranking 14

2.3 Classification-Based Label Ranking 17
2.3.1 Pairwise Label Ranking 18
2.3.2 Probability-Based label ranking 19
2.3.3 Label Ranking based on the Plackett-Luce Model . . . 19
2.4 Classification by Ranking Reduction 20
2.5 Artificial Neural Network (ANN) 21
2.5.1 Feedforward *(FF)* ANN 22
2.5.2 Back Propagation (*BP*) 22
2.5.3 Activation functions 23
2.5.4 Dropout Regularization 23
2.5.5 *ANN* for Ranking 24
2.6 Deep Learning 24
2.6.1 Convolution Neural Network (*CNN*) 25
2.6.2 Pooling Layer 28
2.7 Human Visualization and Frame Variation 28
2.8 Subgroup Discovery (*SD*) 28
2.8.1 *SD* Methodology 29
2.9 Potential Applications 30
2.9.1 *PNN* Applications 30
2.9.2 *PN* Applications 31
2.9.3 Subgroup Preference Neural Network*SGPNN* Applications 31

3 Preference Neural Network **33**
3.1 Introduction 33
3.2 Problem Formulation 35
3.3 *PNN* Components 36
3.3.1 Activation Functions 36
3.3.2 Ranking Loss Function 38

3.3.3 *PNN* Structure 40
3.3.4 Baseline Algorithm 44
3.4 Network Evaluation 44
3.4.1 Activation Functions Evaluation 45
3.5 Experiments 49
3.5.1 Data sets 50
3.5.2 Results 51
3.5.3 Computational Platform 57
3.5.4 Discussion 57
3.6 Conclusion 58

4 Preference Net : Classification by Ranking Reduction 59
4.1 Introduction 59
4.2 *PN* Components 62
4.2.1 Multiple *PNN* 62
4.2.2 Preference Neuron 62
4.2.3 Preference Neural Network 62
4.3 *PN* Structure and Processing 62
4.3.1 Image Prepossessing 64
4.3.2 Feature Extraction 65
4.3.3 Network Structure 67
4.4 Algorithms 69
4.4.1 Kernel size 69
4.4.2 Baseline Algorithm 69
4.4.3 Complexity Analysis 69
4.4.4 Choosing The Kernels 71
4.4.5 Color Image Compensation 72
4.5 Network Evaluation 72
4.5.1 Multiple Kernel Evaluation 73

4.5.2 Dropout Regularization 73
4.6 Experiments 75
4.6.1 Data sets 75
4.6.2 Results 76
4.6.3 Discussion and Future Work 79
4.7 Conclusion 80

5 Subgroup Preference Neural Network 81
5.1 Introduction 81
5.2 Subgroup Preference Neural Network (*SGPNN*) 83
5.2.1 Multi Activation Function Neuron (*MAFN*) 85
5.2.2 *SGPNN* Architecture 86
5.3 Data Preparation and Learning Algorithm 87
5.3.1 Conjoint Data 87
5.3.2 Ranking Unification 88
5.3.3 ***SGPNN* Learning Steps** 89
5.3.4 Dropout Regularization 92
5.4 Experiments 92
5.4.1 Datasets 92
5.5 SGPNN Results 96
5.5.1 Relevant Subgroup Data 97
5.5.2 Non-Relevant Subgroup Data 97
5.6 Discussion 99
5.6.1 Convergence Fluctuation 99
5.7 Conclusions and Future Works 100

6 Preference networks in BCI Application 103
6.1 Introduction 103
6.2 Material and Methods 106

6.2.1 Converting Classification into Ranking 106
6.3 Data Preparation 106
6.3.1 SGPN to solve the problem of class overlap 106
6.3.2 ***SGPNN*** **Learning Steps** 109
6.3.3 Architecture and Data Processing 110
6.4 *MI-EEG* Dataset 110
6.4.1 Dataset 2B 111
6.5 Experiment Flow 111
6.6 Classification Results 112
6.7 Discussion 112
6.8 Conclusion 113

7 Conclusion and Future Work **115**

A **119**
A.1 *PNN* Proof of Convergence 119
A.2 **Supplemental Material** 122
A.3 Video Files 122
A.4 Dataset Files 122
A.4.1 Synthesized data used for SGPNN 122

B **123**
B.1 Symbols and Abbreviations 123

C **125**
C.1 **Video Snapshoot of *PNN* Ranking Convergence.** 125

D **133**
D.1 **Snapshoots of *SGPNN* Ranking Convergence.** 133

Bibliography **145**

List of Figures

Figure **Page**

1.1 Research Scope 3

1.2 Subgroup, incomparable $\perp$ and indifference $\sim$ are New types of labels relations 5

1.3 The thesis structure 10

2.1 Preference Learning Domain 12

2.2 Preference Learning (*PL*) Methods 13

2.3 Preference Relations 14

2.4 *LPC* approach 19

2.5 SD Methodology 29

3.1 Ranker *NN* Sample output image of video file demonstrates the evaluation of *NN* using Sigmoid Activation for ranking. . . . 34

3.2 *PSS* activation function where $n = 3$ and step width $w = 1$ . . . 37

3.3 *SS* activation function where $n = 6$ and 9 and boundary $b = 1$ and 10 in (a) and (b) respectively. 38

3.4 Ranker network and *PNN* evaluation in terms of *RMS* and *spearman* correlation error functions using iris DS in (a) and (b) respectively. 39

3.5 The structure used in both ranker *ANN* and *PNN* where $\varphi_{n=3}$, $f_{in} = 3$ and $\lambda_{out} = 3$, per $\langle x_1, \pi_1 \rangle$, $\mathcal{L} \in \{\lambda_a, \lambda_b, \lambda_c\}$ where $\pi_1 = \{1,2,3\}$. and comparison of the convergence for both *NN*'s. The demo video of convergence of two *NN* in the link Elgharabawy (2020*c*). 40

3.6 Multiple layer label ranking comparison of benchmark data sets Cheng et al. (2009) results using the *PNN* and *SS* functions after 100 epochs and learning rate = 0.007. 41

3.7 The structure of preference neuron where $\varphi_{n=4}$. 41

3.8 *PNN* where $\varphi_{n=16}$, $f_{in} = 16$ and $\lambda_{out} = 16$, per $\langle x_1, \pi_1 \rangle$, $\mathcal{L} \in \{\lambda_a, \lambda_b, \lambda_c, \lambda_d\}$ where $\pi_1 = \{1,2,3,4,\ldots,16\}$. 43

3.9 *PNN* activation function comparison using complete labels and 60% missing labels in (a) and (b), respectively. 46

3.10 The graphical comparison between convergence of *Sigmoid* and *SS* functions to rank stock dataset, (**a**) *Sigmoid* has $\tau = 0.3597$ and epoch = 200. (**b**) *Sigmoid* has $\tau = 0.7876$ and epoch = 1600. (**c**) *SS* has $\tau = 0.4975$ and epoch = 30. (**d**) *SS* has $\tau = 0.8147$ and epoch = 700. 47

3.11 ROC of three label ranking on the wine data set using *PNN* h.n=100 and 50 epochs 48

3.12 The confusion matrix of testing data for wine, glass DS where τ = 0.947, 0.84, Accuracy = 0.935 and 0.8 in (a) and (b) respectively. 49

3.13 Training and validation performance without and with dropout regularization approach in (a) and (b) respectively. 50

3.14 Ranking performance comparison of *PNN* with other approaches. 53

4.1 Classification by ranking reduction flow 63

4.2 Image pixel ranking for each flattened window. 65

4.3 Sample of moving objects in (a) and (b) without and with averaging by window 2x2. The ranking of two flattened images are $\rho = 0.216$ and 0.79 in (a) and (b), respectively. Sample of moving noisy object in (c) and (d) without and with image averaging by a window of 2x2. The ranking of two flattened images are $\rho = 0.137$ and 0.75 in (c) and (d) respectively. and $\rho = -0.18$ and 0.71 of scaled circle in (e) and (f), respectively. 66

4.4 Detecting the similarity in remote sensing and face recognition by ranking the image pixels after averaging the pixels using a 2x2 window. 67

4.5 The *PN* structure has three kernels and three *PNN*s where $\varphi_{n=2}$, $f_{1in} = 16, f_{2in} = 81, f_{3in} = 169$ and $\lambda_{out} = 15$, per $\langle x_1, \pi_1 \rangle, \pi \in \{\lambda_1, \lambda_2, \lambda_3 \cdots, \lambda_{15}\}$. 72

4.6 Image averaging using image width window size applied in MNIST , MNIST-Fashion and Cifar-10 datasets in (a),(c), and (e) and the images after averaging using 2X2 in (b), (d), (f) in order to choose the best kernel windows by searching for the best correlation for all window sizes 73

4.7 Comparison between the number of ranker kernels used in PN; one kernel (15X15) , two kernels (15X15 and 20X20), three kernels (15X15, 20X20 and 25X25),four kernels (10X10, 15X15, 20X20 and 25X25), five kernels (5X5, 10X10, 15X15, 20X20 and 25X25) for training 10 images of Mnist dataset where image size is 28X28. as shown by increasing number of kernels the performance is reached to a stable convergence rate doesn't reach faultless ranking $\rho = 1.0$ as mentioned by Ailon Ailon & Mohri (2007). 74

4.8 weighted ensemble average models flow of the image classification using *PN*s parallel processing for each chunk of training data. 77
4.9 *ROC* evaluation of 500 images of Mnist in the first 50 epochs using *PN* output ranking in (a). The output ranking after ranking score applied for ranking reduction in (b) where ranking 1 and 2 using the score and inverse of score function $f_{s2} = Max(\lambda)$ and $f_{s1} = notMax(\lambda)$ 78

5.1 An example of two subgroups architecture of *SGPNN* to rank conjoint data from two subgroups data, each group has 4 and 3 labels respectively, where $\varphi_{1_{n=4}}$, $\varphi_{2_{n=1}}$, $f_{in} = 4$. A video demo of 2-subgroup architecture is available in (Elgharabawy 2020*c*) . . 84
5.2 The structure of the *MAFN* where $\varphi_{1_{|n=4}}$ and $\varphi_{2_{|n=3}}$. 85
5.3 Two subgroups architecture of *SGPNN*, each group has 4 and 3 labels respectively, where $\varphi_{1_{n=4}}$, $\varphi_{2_{n=1}}$, $f_{in} = 4$. A video demo of 2-subgroup architecture is available in (Elgharabawy 2020*c*) 87
5.4 Three subgroups architecture *SGPNN* used in ranking emotions dataset where $\varphi 1_{n=4}, \varphi 2_{n=3}, \varphi 3_{n=3}$, and $f_{in} = 4$. the second subgroup is represented by one node that has 3 values (1, 2, and 3) mapped to preference relations $\lambda_e = \succ, \sim, \prec$. 90
5.5 Training convergence of german election and iris,wine and stock using *PNN* and *SGPNN* in (a) and (b) respectively. 97
5.6 video image of *SGPNN* convergence using toy dataset 101

6.1 the MI-EEG has BCI competion IV 2B has two problems, class overlap, unbalance data and both in (a), (b), (c) respectively. . . 105
6.2 Matrix plot of the dataset 2b where data are unbalanced and separability is low 107

6.3 Example of class overlapping unification in (a), Using *SGPNN* and *PNN* to learn the classoverlap and separable data in (b) . . 109
6.4 SGPNN architecture used for classification BCI competition VI dataset 2B , each group has 2 labels represents the opposite class classification, where $\varphi_{1_n=2}$, $\varphi_{2_n=2}$, $f_{in}=6$. 110

C.1 *PNN* at epoch 34 125
C.2 *PNN* at epoch 37 126
C.3 *PNN* at epoch 39 127
C.4 *PNN* at epoch 40 128
C.5 *PNN* at epoch 44 129
C.6 *PNN* at epoch 45 130
C.7 *PNN* at epoch 48 131
C.8 *PNN* at epoch 62 132

D.1 *SGPNN* at epoch 1 133
D.2 *SGPNN* at epoch 2 134
D.3 *SGPNN* at epoch 10 135
D.4 *SGPNN* at epoch 11 136
D.5 *SGPNN* at epoch 20 137
D.6 *SGPNN* at epoch 30 138
D.7 *SGPNN* at epoch 40 139
D.8 *SGPNN* at epoch 60 140
D.9 *SGPNN* at epoch 70 141
D.10 *SGPNN* at epoch 90 142
D.11 *SGPNN* at epoch 97 143
D.12 *SGPNN* at epoch 125 144

List of Tables

Table **Page**

2.1 Performance comparison of the recent architectures of different categories. Top 5 error rate is reported for all architectures. . . 27

3.1 *ANN* types used in initial experiment. 44

3.2 Benchmark data sets for label ranking; preference mining Cláudio (2018), real-world data sets Grbovic et al. (2013) and semi-synthetic (*s*-*s*) Cheng et al. (2009). 51

3.3 Benchmark data sets for label ranking; preference mining Cláudio (2018), semi-synthetic (*SS*) Cheng et al. (2009) and real-world data sets 51

3.4 *PNN* performance comparison with various approaches: supervised clustering (Grbovic et al. 2013), supervised decision tree (Cheng et al. 2009), *MLP* label ranking (Ribeiro et al. 2012), Kernel Ridge Regression (K.R.R)(Korba et al. 2018) and label ranking tree forest (*LRT*) (de Sá et al. 2017) 54

3.5 Preference mining ranking performance in terms of the Kendall τ coefficient and learning step and number of hidden neurons. . 55

3.6 Comparison between *PNN* and supervised clustered on biological real world data in terms of $Loss_{LR}$ 55

3.7 *PNN* ranking performance in terms of τ coefficient, learning step and number of hidden neurons. 56

4.1 Comparison between *CNN* and *PN*. 69
4.2 The number of window correlated of two images in Fig. 4.6 (a) of class 3 of *Mnist* dataset by using image averaging 2x2 and 4x4 window using different window size from 3x3 to 28x28. . . 71
4.3 *PN* results on datasets(Mnist, Mnist-Fashion, and CIFAR-10) where The accuracy and sensitivity before and after applying score function. 77
4.4 Comparison of classification on CIFAR-10 Krizhevsky (2009*a*) and Fashion-Mnist data set Xiao et al. (2017*a*) Data sets using different convolution models 79

5.1 Datasets (rest-food-services (Vargas-Govea et al. 2011),german-2005/9 (Rebelo 2018*a*,*b*),emotions (Trohidis & Tsoumakas 2011), sushi (de Sá 2018), iris, wine, and stock (Cheng et al. 2009)) used for *SGPNN* evaluation. 93
5.2 Positive and negative Emotional feelings subgroups relations according to the ranking of the labels for each group. labels per each group are extracted from emotion dataset (Trohidis & Tsoumakas 2011). 95
5.3 *SGPNN* and *PNN* ranking testing data performance comparison of subgroup datasets. 98
5.4 *SGPNN* and *PNN* performance comparison with state-of-the-art approaches: supervised clustering (Grbovic et al. 2013), supervised decision tree (Cheng et al. 2009), multi-layer perceptron label ranking (Ribeiro et al. 2012), and label ranking tree forest (*LRT*) (de Sá et al. 2017). 98

6.1 subjects correlation ρ and accuracy using *PN EEG* classification *BCI* Competition *IV* Dataset *2B* 113

6.2 subjects correlation ρ and accuracy using *PN EEG* classification *BCI* Competition *IV* Dataset *2B* and data separability and duplication θ 113