

Thesis for the Degree of Ph. D.

Deep Neuro-Fuzzy Networks with interpretability for classification

School of Electronics Engineering, Major in Signal Processing
The Graduate School

Nguyen Tuan Linh

June 2020

The Graduate School
Kyungpook National University

Deep Neuro-Fuzzy Networks with interpretability for classification

Nguyen Tuan Linh

School of Electronics Engineering, Major in Signal Processing
The Graduate School

Supervised by Professor Gin-jin Jang
Co-supervised by Professor Minhoo Lee

Approved as a qualified thesis of Nguyen Tuan Linh
for the degree of Ph. D. by the Evaluation Committee

June 2020

Chairperson

Sangmoon Lee

Minho Lee

Gil-Jin Jang

Hoyoung Jung

Sungmoon Jeong

The Graduate School Council
Kyungpook National University

Contents

I.	Introduction.....	1
II.	Related works	13
III.	Deep Convolutional Neuro-Fuzzy Network.....	21
3.1	Convolutional Neuro-Fuzzy Network	21
3.1.1	The proposed CNFN Model	21
3.1.2	CNFN training.....	24
3.1.3	CNFN architecture for text classification.....	26
3.2	Multimodal Convolutional Neuro-Fuzzy Network	29
3.2.1	CNFN for audio feature extraction.....	30
3.2.2	CNFN for text feature extraction.....	31
3.2.3	CNFN for visual feature extraction	34
3.2.4	Feature set visualization	36
3.2.5	Interpretable feature selection by recursive feature elimination and causality analysis.....	37
IV.	Attentive Hierarchical ANFIS	39
4.1	Introduction	39
4.2	Attentive ANFIS (A-ANFIS)	40
4.3	Attentive Hierarchical ANFIS	44
4.3.1	Attentive unit selector	44
4.3.2	ANFIS classifier	46
V.	Attentive Convolutional ANFIS.....	49
5.1	Introduction	49
5.2	Optimal input feature subsets by evolutionary algorithm.....	49

5.3	AConvANFIS	51
5.4	AConvANFIS training.....	53
VI.	Experiments.....	55
6.1	Sentiment analysis with Convolutional Fuzzy-Neural Network	55
6.1.1	Model configuration	55
6.1.2	Dataset and preprocessing.....	57
6.1.3	Results and discussion.....	59
6.1.4	Feature set visualization	62
6.2	Emotion classification of movie clips with Multi-modal Convolutional Fuzzy-Neural Network	67
6.2.1	Unimodal emotion understanding	69
6.2.2	Multimodal emotion understanding	75
6.3	Cancer diagnostic with AH-ANFIS and AConvANFIS	79
6.3.1	Colorectal cancer recurrence prediction.....	79
6.3.2	Breast cancer diagnostic.....	84
VII.	Interpretability Analysis	88
7.1	Interpretable AI by feature and fuzzy rule analysis.....	88
7.2	Activated rules extraction.....	95
7.3	Critical rules selection by recursive rule elimination	97
VIII.	Conclusion and future works	101
	Reference.....	105

List of Figures

Figure 3.1.	A conceptual framework for Convolutional Neuro-Fuzzy Network (CNFN).....	21
Figure 3.2.	CNFN for text sentiment analysis.	26
Figure 3.3.	Multimodal sentiment analysis framework for movies	29
Figure 4.1.	A conceptual framework of the proposed AH-ANFIS model.	40
Figure 4.2.	A-ANFIS with attentive rule selector.	41
Figure 4.3.	Structure of attentive A-ANFIS units selector.....	45
Figure 4.4.	ANFIS classifier.	46
Figure 5.1.	A conceptual framework of the proposed AConvANFIS model.....	51
Figure 5.2.	ANFIS classifier with multiple consequence unit and softmax layer.	53
Figure 6.1.	Projection of scatter plots of test input samples	62
Figure 6.2.	Projection scatter plots of output feature set extracted by convolutional layers	64
Figure 6.3.	Projection scatter plots of feature set extracted by convolutional stage .	66
Figure 6.4.	Distribution of centers of defuzzification membership function at initial (a) and after model trained (b)	67
Figure 6.5.	Projection of scatter plots of audio test input samples.....	69
Figure 6.6.	Projection scatter plots of audio set extracted by convolutional stage....	70
Figure 6.7.	Visual critical features selection by RFE.....	74
Figure 6.8.	Result of evolutionary algorithm for permutation selection.	80
Figure 7.1.	Critical features selected video modality.....	89
Figure 7.2.	An example of audio feature.	89
Figure 7.3.	Critical features selected from audio modality.	91
Figure 7.4.	Critical features selected of text modality.	91
Figure 7.5.	Examples of input sentences with emotion words extraction.	92
Figure 7.6.	ANFIS rule set visualization.....	92
Figure 7.7.	Selection of rules for interpretability.....	96
Figure 7.8.	An example of extracted rule from AH-ANFIS for CRC model.....	96
Figure 7.9.	Critical rule sets selected UCI breast cancer dataset.	99

List of Tables

Table 3.1.	CNFN model parameters for audio feature extraction.....	31
Table 3.2.	CNN and CNFN model parameters for text emotion understanding	32
Table 3.3.	CNFN model parameters for video emotion understanding	35
Table 6.1.	CNN and CNFN model parameters for text sentiment analysis.	56
Table 6.2.	Summary statistic of used datasets	57
Table 6.3.	Some samples of sentences in MR dataset	58
Table 6.4.	Comparison of classification accuracy of CNN and CNFN for MR dataset using cross-validation.....	59
Table 6.5.	Summary of classification accuracy of CNN, CNFN, and CNFN w/o FuzzConv for sentiment analysis	60
Table 6.6.	Comparison performance reduced by adding noise to MR dataset.....	60
Table 6.7.	Some samples of ambiguity sentences in MR dataset	61
Table 6.8.	Comparison of Silhouette score.....	66
Table 6.9.	Comparison of average classification accuracy of CNN and CNFN for audio feature	71
Table 6.10.	Comparison of Silhouette score.....	71
Table 6.11.	Comparison of average classification accuracy of CNN and CNFN for text	72
Table 6.12.	Comparison of training and testing time.....	73
Table 6.13.	Comparison of average classification accuracy of CNN and CNFN for video modality.....	73
Table 6.14.	Feature selection result	75
Table 6.15.	Comparison of classification accuracy of M-CNN and M-CNFN.....	76
Table 6.16.	Examples of ambiguous inputs.	77
Table 6.17.	CRC variable permutation selected by evolutionary algorithm.	81
Table 6.18.	AH-ANFIS model hyper-parameters for CRC recurrence prediction	81
Table 6.19.	CNN and AH-ANFIS model configurations for CRC recurrence prediction	82
Table 6.20.	CNN and AConvANFIS model parameters for CRC recurrence prediction	83
Table 6.21.	Comparison of average classification F-score of SVM, ANFIS, CNN, AH-ANFIS, and AConvANFIS for CRC dataset.	84
Table 6.22.	Wisconsin diagnostic breast cancer dataset description	84
Table 6.23.	CNN and AH-ANFIS model configurations for breast cancer diagnostic.....	85

Table 6.24.	AH-ANFIS model hyper-parameters for breast cancer diagnostic	85
Table 6.25.	Breast cancer dataset variable permutation optimized by evolutionary algorithm.....	86
Table 6.26.	CNN and AConvANFIS model parameters for breast cancer diagnostic	86
Table 6.27.	Comparison of average classification F-score of SVM, ANFIS, CNN, AH-ANFIS, and AConvANFIS for breast cancer diagnostic dataset	87
Table 7.1.	Critical analysis of CRC input features result.....	98
Table 7.2.	Critical analysis of breast cancer input features result.....	98
Table 7.3.	Recursive rule elimination result for breast cancer dataset.....	99
Table 8.1.	Summary of proposed models	103

I. INTRODUCTION

Deep Learning (DL) has emerged as a family of powerful machine learning models with superior classification performance in AI applications to improve diagnosis [1], classification, and prediction of clinical outcome [2]. This can be attributed to the deep hierarchical structure that can effectively capture relevant high-level abstractions and characterize training data very well in a layer-by-layer manner [3]. It has been mentioned that deep neural networks are forming an efficient internal representation of the learning problem. Still, it is unclear how this competent representation is distributed layer-wise and how it arises from learning [4]. This lack of transparency in the training process often causes crucial trust-related problems in critical application areas such as health care where validation is essential. A vital component of an AI system is the ability to explain the decisions made by it and the process through which they are made. These explanations offer an insight into why a particular action has been chosen.

Convolutional Neural Networks (CNNs) are amongst the most prevalent architectures for deep learning (DL), that empower big data feature extraction with robustness and accurateness. They effectively draw out from low-level input data to high-level abstraction features due to the benefit of a massive number of samples. However, due to inadequate information or complexity in the input feature, data may be ambiguous or vague which is mostly considered

as data ambiguity [5]. Performance of CNNs in emotion understanding from video clips which have essential syntactic, semantic, and visual ambiguity is insufficient. CNN is a totally deterministic system used in a “black-box” behavior that impossible to manipulate data ambiguity [6].

Fuzzy inference system (FIS) is an effective mechanism for modeling human perception and reasoning [7]. The mathematical framework for ambiguous data processing may be provided by the possibility theory of fuzzy logic. Numerical computations performed by fuzzy logic using linguistic labels and fuzzy degrees of membership, which are represented as degrees of truth [6]. Humans could easily interpret the feature extraction and the reasoning process from fuzzy rules and fuzzy inference. Nevertheless, fuzzy rules are needed to determine by human experts, and the learning capability of fuzzy systems is deficient. By incorporating fuzzy logic with neural network, neuro-fuzzy networks can automatically learn the fuzzy membership functions [8]. Therefore, the fuzzy system parameter could be obtained from a large volume of training data.

Today, throughout the era of the Internet, and with the explosion of social media, it is imperative to dig into key and relevant knowledge from the multitude of data available in it. These usually come in the form of text and express the reader's love for content such as goods, utilities, books, hotels, etc. Text is a good source for sharing your opinions, emotions, and feelings. Languages are not only used for communication, but they also convey the