Master program: ***Advanced Computing Systems***

NEUROPROCESSING

Projects Description /Contents 2018-2019

Project 1. Character Recognition using SOM and Multi-Layer Feed-Forward Network

Description:

Develop an application that allows you to load different characters from image files and automatically recognize the input character.
Recommended steps:
Step 1 – Represent the input characters as matrices in a binary format.
Step 2 – Extract the principal characteristics and group the characters in clusters (SOM stage).
Step 3 – For each group train a dedicated back-propagation network.
Step 4 – Implement a testing part, where new characters are presented to the system and analyze the output of the system. Evaluate (compute) the recognition accuracy.

Issues that must be addressed for presentation of the project:

A. Multilayer Feed-forward Networks:

- The Feed-forward Network Architecture
- The Standard Back-propagation Algorithm
  - Learning Factors:
  - Initial Weights
  - Cumulative Weight Adjustments versus Incremental Updating
  - Steepness of the Activation Function
  - Learning Constant
  - Avoiding Local Minima
  - Momentum Method
  - Choosing Right Network Architecture and Number of Hidden Neurons

B. SOM (*Self-Organizing Maps*):

- The SOM Architecture and the SOM Algorithm
- Improvements:
  - Initialization of the SOM Algorithm
  - "Optimal" Learning Rate Factor
  - Effect of the Form of Neighborhood Function
- Clustering with Self-Organizing Networks

C. **Hierarchical neural system: SOM + Multilayer Feed-forward Networks**

- Building the hierarchical neural system
- Using the hierarchical system in Character Recognition (approach and implementation)
- Performance evaluation (recognition accuracy)
- Concluding Remarks

Project 2. Neural System based on Hopfield Network for automatic handwritten numbers recognition

Description:

Develop an application that allows you to draw into a grid with size of 20*20 different numbers. The application also will implement the Hopfield algorithm to train with different numbers drawn in the grid and automatically recognition the numbers, even if there are in testing part some mistakes in drawing number.

Issues that must be addressed for presentation to the class of the project:

A. Hopfield Model as Recurrent Network:

- Mathematical Foundation
- Convergence to Stable States
- Learning in Hopfield Networks
- Hopfield Networks in Applications

B. Hopfield Neural Network Implementation:

- Implementing the Hopfield Network in software
- Using the Hopfield Network in handwritten numbers recognition
- Performance evaluation (recognition accuracy)
- Concluding Remarks

Bibliography:

1. http://en.wikipedia.org/wiki/Hopfield_net
2. http://facstaff.cbu.edu/~pong/ai/hopfield/hopfieldapplet.html (requires running of a java applet)

Project 3. Text (WEB) Document clustering using Self-Organizing Network

Description:

Develop an application that implement the SOM (*Self-Organizing Map*) neural network clustering algorithm for automatic text document grouping. For this subject the student will receive preprocessed text files (Reuters database) where documents are represented as sparse words vectors. Also need to be implemented the testing part of learning algorithm to evaluate the performance of the algorithm (use measure like accuracy, compactness, separability, balance, …).

Issues that must be addressed for presentation to the class of the project:

A. SOM theory:

- Biological Background
- The SOM Architecture and the SOM Algorithm
- Improvements:
  - Initialization of the SOM Algorithm
  - "Optimal" Learning Rate Factor
  - Effect of the Neighborhood Function
  - Practical Advices for Construction of a Good Maps

B. Self-Organizing Network implementation:

- Documents clustering with Self-Organizing Networks, approach and implementation
- Performance evaluation (recognition accuracy)
- Concluding Remarks

Bibliography:

1. Jiawei Han, Data Mining: Concepts and Techniques, Morgan Kaufmann publisher, 2006
2. http://en.wikipedia.org/wiki/Self-organizing_map
3. http://en.wikipedia.org/wiki/Data_mining
4. https://pdfs.semanticscholar.org/efff/7c103815fe5400c17049958d7bb9f9f68b51.pdf
5. Teuvo Kohonen - WEBSOM - Self-organizing maps of document collections

Project 4. Handwritten Digit Recognition using PCA and Back-Propagation Network

Description:

Develop an application that allows you to load different characters from image files and automatically recognize the input character.

Recommended steps:
Step 1 – Represent the input character as a matrix in a binary format
Step 2 – Extract the principal characteristics and group the characters based on their representation, using PCA (Principal Component Analysis)
Step 3 - Train a back-propagation network.
Step 4 – Implement a testing part, where new characters are presented to the network and analyze the output of the network. Evaluate (compute) the recognition accuracy.

Issues that must be addressed for presentation to the class of the project:

A. Multilayer Feed-forward Networks:

- The Feed-forward Network Architecture
- The Standard Back-propagation Algorithm
  - Learning Factors:
  - Initial Weights
  - Cumulative Weight Adjustments versus Incremental Updating
  - Steepness of the Activation Function
  - Learning Constant
  - Avoiding Local Minima
  - Momentum Method
  - Choosing Right Network Architecture and Number of Hidden Neurons

B. PCA (*Principal Component Analysis*):

- PCA theory, PCA as method to reduce the input vector dimension (data compression)
- PCA implementation

C. Recognition system implementation:

- Neural system for digit recognition (approach and implementation); software application design
- Performance evaluation (recognition accuracy)
- Concluding Remarks

Bibliography:

1. Industrial Applications of Neural Networks pages – 162 to 189
2. Tom Mitchell, Machine Learning, Publisher McGraw-Hill Science, **ISBN:** 0070428077, 1997

Project 5. Neuro-fuzzy system for Geometric Shape recognition: on-line shape recognition from hand drawn shapes

Description:

Develop an application that allow you to draw by hand simple geometric shapes (rectangle, circle, and triangle) and automatically recognize the shape and repaint it.
Recommended steps:

Step 1 - Feature extraction – extracting the significant points

i. Resampling (computing the center of the shape)
ii. Aim to eliminate crossover and obtain sample points

Step 2 - Building tangent vectors along shape Boundary
Step 3 – Shape classification by a **Binary Synaptic Weights** network

Issues that must be addressed for presentation to the class of the project:

A. The BSW Algorithm:

- Mathematical Foundations
- Advantages in Implementation

B. **Geometric Shape Recognition with Fuzzy Filtering and Neural Network Classification:**

- Feature Extraction
- Fuzzy Function Filtering
- Shape Classification and Training by a BSW Network
- Alternative Solutions for Shape Classification Problem

C. Recognition system implementation:

- Software application: approach and implementation
- Performance evaluation (recognition accuracy)
- Concluding remarks

Bibliography:

1. Industrial Applications of Neural Networks pages – 3 to 31
2. http://en.wikipedia.org/wiki/Binary_neuron
3. http://ukpmc.ac.uk/articlerender.cgi?artid=1168438

Project 6. Text (WEB) Document classification using Genetic Algorithm

Description:

Develop an application that implement a Genetic Algorithm that using a feed-forward neural network to learn the text document classification problem. For this subject the students will receive preprocessing text files (Reuters database) where documents are represented as sparse words vector into term frequency vector. Also need to be implemented the testing part for evaluate the performance of the implemented algorithm (use measure like accuracy, precision, recall, true positive rate,…).

Issues that must be addressed for presentation to the class of the project:

- The Genetic Algorithm:
  - Chromosome representation
  - Operators (selection, mutation, crossover)
  - The fitness function
- The Feed-forward Network Architecture
  - Learning Factors:
  - Activation Function
  - Avoiding Local Minima
  - Choosing Right Network Architecture and Number of Hidden Neurons
- Methods for evaluating the algorithm.
- Concluding Remarks

Bibliography:

1. Jiawei Han, Data Mining: Concepts and Techniques, Morgan Kaufmann publisher, 2006
2. http://en.wikipedia.org/wiki/Data_mining
3. http://en.wikipedia.org/wiki/Genetic_algorithm
4. Tom Mitchell, Machine Learning, Publisher McGraw-Hill Science, ISBN: 0070428077, 1997

Project 7. Text (WEB) Document classification using LVQ algorithm

Description:

Develop an application that implement the LVQ (Learning Vector Quantization) classifier algorithm for text document classification. For this subject the student will receive preprocessed text files (Reuters database) where documents are represented as sparse words vector into term frequency vector. Also need to be implemented the testing part for evaluate the performance of the implemented algorithm (use measure like accuracy, precision, recall, true positive rate,…).

Issues that must be addressed for presentation to the class of the project:

- The LVQ architecture and algorithm, initializing the weights and problems of the algorithm.
- Improvements
  - LVQ 2 modification.
  - LVQ 2.1 modification.
  - LVQ 3 modification.
- Document classification using LVQ – approach and implementation
- Methods for evaluating the algorithm.
- Concluding Remarks

Bibliography:

1. Jiawei Han, Data Mining: Concepts and Techniques, Morgan Kaufmann publisher, 2006
2. http://en.wikipedia.org/wiki/Data_mining
3. http://ccy.dd.ncu.edu.tw/~chen/course/Neural/ch4/index.htm
4. http://www.cs.bham.ac.uk/~pxt/NC/lvq_jb.pdf

Project 8. Text (WEB) Document classification using Naïve Bayes algorithm

Description:

Develop an application that implement the Naïve Bayes classifier algorithm for text document classification. For this subject the student will receive preprocessed text files (Reuters database) where documents are represented as sparse words vector. Also need to be implemented the testing part for evaluate the performance of the algorithm (use measure like accuracy, precision, recall, true positive rate,…).

Issues that must be addressed for presentation to the class of the project:

- Mathematical background for Bayes Theory
- The Naïve Bayes algorithm. Bayesian classification.
- Laplace smoothing.
- Modification of the Naïve Bayes algorithm for multi-class classification problem.
- Document classification using Naïve Bayes – approach and implementation.
- Evaluating the algorithm
- Concluding Remarks

Bibliography:

1. Jiawei Han, Data Mining: Concepts and Techniques, Morgan Kaufmann publisher, 2006
2. Tom Mitchell, Machine Learning, Publisher McGraw-Hill Science, ISBN: 0070428077, 1997
3. http://en.wikipedia.org/wiki/Data_mining
4. http://en.wikipedia.org/wiki/Naive_Bayes_classifier

Project 9. Text (WEB) Document classification using Decision Trees algorithm

Description:

Develop an application that implement the Decision Trees classifier algorithm for text document classification. For this subject the student will receive preprocessed text files (Reuters database) where documents are represented as sparse words vector. Also need to be implemented the testing part for evaluate the performance of the algorithm (use measure as accuracy, precision, recall, true negative rate,…).

Issues that must be addressed for presentation to the class of the project:

- Decision Tree algorithm. General approach.
- Methods for selecting the best nodes. Mathematical Foundation.
- Methods for pruning the tree.
- Methods for algorithm evaluating.
- Document classification using the decision tree - approach and implementation.
- Concluding Remarks

Bibliography:

1. Tom Mitchell, Machine Learning, Publisher McGraw-Hill Science, **ISBN:** 0070428077, 1997
2. http://en.wikipedia.org/wiki/Decision_tree
3. http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/mlbook/ch3.pdf

Project 10. Text (WEB) Document classification using SVM algorithm

Description:

Develop an application that implement the Support Vector Machine classifier algorithm for text document classification. For this subject the student will receive preprocessed text files (Reuters database) where documents are represented as sparse words vector. Also need to be implemented the testing part for evaluate the performance of the algorithm (use measure as accuracy, precision, recall, true negative rate,…).

Issues that must be addressed for presentation to the class of the project:

- Mathematical background of the kernel based algorithm and the SVM algorithm.
- Selecting the best separating hyperplane.
- Dual optimization problem and the kernel trick.
- Soft margin hyperplanes.
- Methods for algorithm evaluating.
- Document classification using the SVM – SMO (Sequential Minimal Optimization) algorithm - approach and implementation.
- Concluding Remarks

Bibliography:

1. Schoslkopf Bernhard, Smola Alexander - Learning with kernels. Support Vector Machine. MIT Press, London, 2002.
2. Vapnik V – The nature of statistical learning Theory, Springer New York, 1995
3. http://en.wikipedia.org/wiki/Support_vector_machine