



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

DEPARTMENT OF COMPUTER SCIENCE

SOFTWARE DEVELOPMENT FOR ALGORITHMIC PROBLEMS

K23 γ

NNs FOR WIND SPEED PREDICTION

STUDENT ID :

sdi1600151

sdi1700115

Contents

1	ABSTRACT	2
2	Directories and Files	2
3	Compilation and Running	3
4	Implementation Details	4
5	Results	4
5.1	NN Representations	4
5.2	New Representations	4
6	Conclusion	5

1 ABSTRACT

In this project, we've evaluated a pre-trained model using the clustering algorithms implemented in our previous project.

2 Directories and Files

Our implementation code is derived in the following directories:

- app: includes utilization functions called from all interfaces and python source code used to evaluate our pre-trained model.
 - includes: includes utilization libraries
 - src: includes the implementation of utilization functions
- config: includes configuration file
- core: includes functions called from all interfaces
 - cluster
 - * clustering: includes cluster's class implementation. Cluster Class combines all 8 algorithms implemented.
 - * initialization: includes the implementation of initialization algorithms.
 - * assignment: includes the implementation of assignment algorithms.
 - * update: includes the implementation of update algorithms.
 - * search: includes the implementation of functions commonly used by different search algorithms in different interfaces.
 - * utils: includes the implementation of interfaces' utilisation.
 - hash: includes hash function's and amplified hash function's implementation
 - metric: includes the implementation of different metrics (ex:DTW,Manhattan, Euclidean)
 - search: includes the implementation of functions commonly used by different search algorithms in different interfaces.
 - utils: includes the implementation of interfaces' utilisation.
- datasets: includes vectors' dataset and query files.

-
- results: includes search algorithms' results.
 - scripts: includes compilation script for each interface.

3 Compilation and Running

Evaluating N2 Neural Network

Navigate to **scripts** file, then run **sh run_py_predict.sh**. Modify compilation scripts in case you want to change arguments.

Argument Parameters:

- -i input file

Evaluating first dense of N2 Neural Network

Navigate to **scripts** file, then run **sh run_py_new_representation.sh**. Modify compilation scripts in case you want to change arguments.

Argument Parameters:

- -i input file

Clustering results

Navigate to **scripts** file, then run **sh run_new_representations.sh** or **sh run_nn_representations.sh** depending on the dataset you want to use. Modify compilation scripts in case you want to change arguments.

Argument Parameters:

- -i input file
- -c configuration file
- -o output file
- -complete
- -init initialization algorithm
- -assign assignment algorithm
- -update update algorithm

4 Implementation Details

General

- Keras loss functions are used for MAE and MSE computation.
- MAPE is computed by replacing each y_{true} value with the mean value of the row - and then computing the mean value of the error between true and predicted data.

5 Results

5.1 NN Representations

- Input
 - dataset: nn_representations.csv
 - initialization: random
 - assignment: lloyd's simplest
 - update: mean
 - max iterations: 30
- Output
 - number of clusters: 4 Silhouette: 0.286989
 - number of clusters: 12 Silhouette: 0.250577

5.2 New Representations

- Input
 - dataset: new_representations.csv
 - initialization: random
 - assignment: lloyd's simplest
 - update: mean
 - max iterations: 30
- Output
 - number of clusters: 4 Silhouette: 0.417065

– number of clusters: 12 Silhouette: 0.263534

6 Conclusion

Getting the output of our Neural Network we get really good clustering results for the 4 seasons, and just good enough for the 12 months. Comparing to the clustering results of our initial vectorized data - it seems that the whole neural network architecture aims to compute new features given the initial inputs and make predictions. These new representations even though they carry the same information with the initial ones, are far more compacted with new features embedded to them. All in all, that's the role of neural networks, to learn new features by their own and help us understand our data in depth. For that reason the clustering results of NN's output vectors are way better compared to the initial ones.