Egyptian Computer Science Journal

Instructions for Preparation of Manuscripts

Papers submitted for possible publication must be presented in English in one original typescript and two identical copies with a maximum of **10 pages** in .doc or .docx file.

These guidelines include complete descriptions of the fonts, spacing, and related information for producing your submission.

1. **Paper Size**: A4
2. **Margins**: Top: 2.5 cm, Bottom: 2.5 cm, Left: 2.5 cm, Right: 2,5 cm, Header: 1,5 cm, Footer: 1,5 cm

   Do NOT Use footers, headers and page numbers.

3. **Title**: The tit le should be centered in 14pt Times New Ro man, boldface, in itially cap italized and paragraph spacing after 18pt.
4. **Authors**: The Authors name should be centered in 12pt Times New Roman and Paragraph spacing after 12 pt.
5. **Affiliation**: The Affiliation should be centered in 10pt Times New Roman and paragraph spacing after 6pt.
6. **Abstract**:

   (a) Use the word **Abstract** as the title, flush left, in 12 pt Times New Ro man, boldface, paragraph spacing before 42pt, paragraph spacing after 6pt, paragraph indentation before is 3.5 cm and paragraph indentation after is 1.5 cm.

   (b) The abstract is to be in 9pt Times New Ro man and justified. Paragraph indentation before is 3.5 cm and paragraph indentation after is 1.5 cm.

7. **Keywords**:

   (a) Use the word **Keywords** as the title, flush left, in 12 pt Times New Ro man, boldface, paragraph spacing after 24pt and paragraph spacing before 6pt.

   (b) The keywords are to be in 9pt Times New Ro man and justified. Paragraph indentation before is 3.5 cm and paragraph indentation after is 1.5 cm.

Note: use two lines 0 height and 16 cm width, one before the word **Abstract** and the other after the word **Keywords**

8. **Main Text**: Type your main text in 10 pt Times New Ro man and justified. A ll paragraphs and subsequent paragraphs should be indented first line 0,5 cm.
9. **First-order Headings** : For examp le, "**1. Heading**", should be 12pt Times New Ro man, boldface, paragraph spacing before 12pt, paragraph spacing after 12pt, flush left and in itially capit alized. Use a period (".") after the heading number, not a colon.
10. **Second-order Headings** : For example, "**1.1 Heading**", should be 11pt Times New Ro man, boldface, initially capitalized, flush left, paragraph spacing after 12pt and paragraph spacing befo re 12pt. Use a period (".") after the heading number.
11. **Third-order Headings** : For example, "**1.1.1 Heading**", should be 10 pt Times New Ro man, boldface, in itially capitalized, flush left, paragraph spacing after 6pt and paragraph spacing before 6pt. Use a period (".") after the heading number.

    Do NOT Use more than three levels of heading.

12. **Figures and Tables** : All figures and tables should have caption. Figure and table captions should be 10pt Times New Ro man. In itially capitalize only the first word of each figure caption and table title. Figures and tables must be numbered separately. For examp le "Figure 1. Text here", "Table 1. Text here". Figure captions are to be centered below the figures with paragraph spacing after 6pt and paragraph spacing before 6pt. Table tit les are to be centered above the tables, paragraph spacing after 6pt and paragraph spacing before 6pt.
13. **References**: References should be 10pt Times New Ro man, justified, indentation h anging 0,5 cm and paragraph spacing after 6pt.
14. Sample of applying the format instructions is show in the following

Sample of Applying the Format Instructions

Unsupervised Artificial Neural Networks For Clustering Of Document Collections

Abdel-Badeeh M. Salem, Mostafa M. Syiam, and Ayad F. Ayad
Computer Science Department, Faculty of Computer & Information Sciences
Ain Shams University, Cairo, Egypt.
Tel. (+202) 6844284, Fax. (+202) 6828298
Email: absalem@asunet.shams.edu.eg , syiam@worldnet.com.eg, Ayad_fa@hotmail.com

Abstract

The Self-Organizing M ap (SOM) has shown to be a stable neural network model for high- dimensional data analysis. However, its applicability is limited by the fact that some knowledge about the data is required to define the size of the network. In this paper the Growing Hierar chical SOM (GHSOM) is proposed. This dynamically growing architecture evolves into a hierarchical structure of self–organizing maps according to the characteristics of input data. Furthermore, each map is expanded until it represents the corresponding subset of the data at specific level. We demonstrate the benefits of this novel model using a real world example from the document -clustering domain. Comparison between both models (SOM & GHSOM) was held to explain the difference and investigate the benefits of using GHSOM .

Keywords: *Neural networks, Self-Organizing Map, Document Clustering*.

1. Introduction

The Self-Organizing Map (SOM) [1] is an artificial neural network model that is well suited for mapping high-dimensional data into a 2-d imensional representation space. The training process is based on weight vector adaptation with respect to the input vectors. The SOM has shown to be a highly effect ive tool for data v isualization in a broad spectrum of application domains [2] . Especially the utilization of the SOM for info rmat ion retrieval purposes in large free-form document collections has gained wide interest in the last few years [3, 4, 5]. The general idea is to display the contents of a document library by representing similar documents in similar regions of the map. One of the d isadvantages of the SOM in such an application area is its fixed size in terms of the number of units and their particu lar arrangement, which has to be defined prior to the start of the training process. Without knowledge of the type and the organization of the documents it is difficult to get satisfying results without multip le train ing runs using different parameter settings, which obviously is ext remely t ime consuming given the high -d imensional data representation. Recently a number of neural network models inspired by the training process of the SOM and having adaptive architectures were proposed [6]. The model being closest to the SOM is the so -called Gro wing Grid [7], where a SOM-like neural network grows dynamically during training. The basic idea is to add rows or columns to the SOM in those areas where the input vectors are not yet represented sufficiently. More precisely, units are added to those regions of the map where large deviat ions between the input vectors and the weight vector of the unit representing these input data are observed. However, this method will p roduce very large maps, which are d ifficu lt to survey and therefore are not that suitable for large document collect ions. Another possibility is to use a hierarchical structure of independent SOMs [8], where for every unit of a map a SOM is added to the next layer. This means that on the first layer of the Hierarchical Feature Map (HFM) we obtain a rather rough representation of the input space but with descending the hierarchy the granularity increases. We believe that such an approach is especially well suited for the representation of the contents of a document collect ion. The reason is that document collections are inherently structured hierarch ically with respect to different s ubject matters. This is essentially the way how conventional libraries are organized for centuries. However, like with the original SOM , the HFM uses a fixed arch itecture with a specified depth of the h ierarchy and predefined size of the various SOMs on each layer. Again, we need profound knowledge of the data in order to define a suitable architecture. In o rder to co mbine the benefits of the neural network models described above we introduce a Growing Hierarchical SOM (GHSOM). This model consists of a hierarchical architecture where each layer is composed of independent SOMs that adjust their size accord ing to the requirements of the input data. The remainder of the paper is organized as fo llo ws. In section 2 we describe the architecture and the training process of the GHSOM. The used data set and preprocessing steps are demonstrated in section 3. The results of experiments

References:

[1] T. Kohonen, “Self-organized formation of topologically correct feature maps,” Biol. Cy bern. vol. 43, 1982, pp. 59–69.

[2] T.Kohonen, “Self-organizing maps” Berlin, Germany: Springer verlage, 1998.

[3] K. Lagus, T. Honkela, Saski, and T. Kohonen, “Self-organizing maps of document collection: A new approach to interactive exploration” In Proc. Int. Conf. on Knowledge Discovery and Data Mining (KDD-96), Portland, OR, vol.36, 1998, pp. 314-322

[4] D. Merkl, “Exploration of text collections with hierarchical feature maps”. In Proc. Int. ACM SIGIR Conf. on Information Retrieval (SIGIR'97), Philadelphia, PA, vol.62, 1997,pp. 412-419

[5] A. Rauber and D. Merkl, “Findin g structure in text archives” In Proc. Europe an Symp. On Artificial Neural Networks (ESANN98), Bruges, Belgium, vol.18, 2000,pp.410-419

[6] B. Fritzke, “Growing self-organizing networks ------Why?” In Proc. Europ Symp on Artificial Neural Networks (ESANN'96), Bruges, Belgium, vol.16,1998,pp.222-230.

[7] B. Frit zke, “Gro wing grid: a self-organizing network with constant neighborhood range and adaptation strength” Neural Processing Letters, 1997.

[8] R. Miikkulainen, “Script recognition with hierarchical feature maps” Connection Science, 2, 1995.

[9] M. Salem, M. Sy iam, and A. F. Ayad, “Improving self-organizing feature map (SOFM) training algorithm using k-means initialization” In Proc. Int. Conf. on Intelligent Eng. Systems INES, IEEE, vol.40,2003,pp.41-46.

[10] M. Porter, “An algorithm for suffix stripping” Program 14(3), pp. 130-137, 1980.

[11] K. Lagus, and S. Kaski, “Keyword selection method for characterizing text document maps” In Proc of ICANN99, Ninth International Conference on Artificial Neural Networks,IEEE,vol 68, 1999,pp.615-623.