A Nonhierarchical Neural Network Approach For Analyzing Textual Data

Brenda L. Battleson, Hao Chen, Carolyn Evans, Joseph Woelfel
Department of Communication, University at Buffalo (SUNY), Buffalo, New York 14260 USA

Introduction

Artificial neural networks excel at recognizing patterns in textual data. Its pattern recognition capability allows a neural network engine to assign weights representing the multiple connections among concepts. These weights can then be used to create dendrograms or hierarchical concepts in a hierarchical clustering method. Unlike the traditional forward feed-back propagation neural networks, (ANN) the output consists of the total signal received by any node activation value of . The signal sent from any node neuron will be the sum of the signals received from all the other nodes, or i.e., the total signal received by any node and strength of the connection between i and j, the sum of the weighted input network representing the multiple connections among concepts. These weights can then be used to create dendrograms or hierarchical concepts in a hierarchical clustering method. Unlike the traditional forward feed-back propagation neural networks, (ANN) the output consists of the total signal received by any node according to Woelfel, the most nonhierarchical analysis of the CATPAC results is an interactive activation and competition network, and any neuron allows the researcher to interact with the neural network to explore all possible meanings of concepts. Thus, in the resulting output a concept may appear in as many clusters as are appropriate. In this study, a large dataset containing multiple newspaper articles, is examined. Hierarchical and nonhierarchical procedures are compared.

Method

Opinions about the terrorist attacks of September 11, 2001, years of particular interest during the five year anniversary of the event in September 2006. To gauge opinion, sentiment, opinion pieces and editorials were selected and retrieved for the month of September 2006. Using a Cox-Deve method the non analyzed using the CATPAC™ and ThoughtView™ program. Output consisted of neural products of terms used to generate an artificial neural network (ANN) with output consisting of a weighted input network (WIN) and the hierarchical clusters identified using CATPAC analysis and a 3-dimensional map. In both dendograms and 3-dimensional coordinates file, the CATPAC™ software was used for the analysis. Unlike the traditional forward feed-back propagation neural networks, (ANN) an interactive activation and competition network, and antecedents can be an input, unlike output neurons. According to Woelfel, the most nonhierarchical representation of ORBANE is Patrick Louwrens’ interpretation of a neural network (Fig. 1A).

Results

NONHIERARCHICAL CLUSTERING: ORBANE allows the researcher to input “activates” the matrix representing a concept of concepts. Multiple cycles allow the “learning” and weight adjustment of associated neurons. A threshold value determined by the researcher determines what associated neurons are activated. The results are new and often very different clusters of concepts.

Conclusion

The human brain is the most sophisticated example of a parallel distributed processing machine. The language used to express human ideas, attitudes and emotions is evidence of this sophistication. Yet our try to analyze language and human communications with bounded linear methods. In this study, ideas generated with regard to a single context experimentally created provided surprisingly good results. We use in the hierarchical clustering (HCR) method. This method allows the researcher to interact with the neural network to explore all possible meanings of concepts. Thus, in the resulting output a concept may appear in as many clusters as are appropriate. In this study, a large dataset containing multiple newspaper articles, is examined. Hierarchical and nonhierarchical procedures are compared.

References


More information

Please contact Brenda Battleson, Ph.D. for information on the software and related procedures used in this study. All terms included in the set of the output can be obtained by contacting the authors.