Introduction

Artificial neural networks excel at recognizing patterns in complex data. Their pattern recognition capability is achieved by assigning weights representing the multiple connections among concepts. These weights can then be used to create dendrograms or other category concepts in a hierarchical manner. However, this approach has its limitations.

Words often have different meanings depending on the context in which they occur. Thus, a hierarchical clustering method is unable to fully capture the nuance of meaning associated with words. For this reason, a nonhierarchical clustering approach is needed. The ANN makes it possible to reorganize the data into a new way that allows for a “brainstorming” and weight adjustment of associated concepts. A threshold value determined by the researcher determines what associated neurons are activated. The results are new and often very different clusters of concepts.

Method

Opinions about the terrorist attacks of September 11, 2001 were of particular interest during the 5 year anniversary of the event. To gauge opinion, editors, opinion pieces and letters to the editors of all U.S. newspapers indexed in the FACTIVA™ database were retrieved for the month of September, 2006.

A 3 x 444 text file was analyzed using the CATPAC™ text analysis program. Output consisted of vocabulary words used to describe the terrorist attacks. The concepts were clustered into hierarchical and nonhierarchical clusters. Figure 1 shows the relationship among the clusters. The results have been condensed to show the associated names. A nonhierarchical approach allows us to see some of those relationships that may not have been statistical “best fits,” but are nonetheless important in finding meaningful text. Consider the concept SEPTELEVENTH which has a clearly defined cluster identified as KABUL in Fig. 3. When it is paired with another term, NEWS, the concepts with which it was originally clustered are not as important. Furthermore, terms with which it was not seemingly related are now much closer. Indeed, SEPTELEVENTH is related to the nonhierarchical concept CLINTON.

Results

Hierarchical Clustering

The results of the CATPAC analysis show definite clusters:

- The researcher is not limited to analyzing one concept at a time. Multiple concepts may be initially activated. (Fig. 5 & 6)
- The disadvantage of hierarchical cluster analysis is that we see only part of the picture. Concepts are placed in one “overall best fit” cluster where it really is in one or many clusters depending on the method used. Hierarchical cluster analysis that many concepts were grouped as one cluster when in reality, they can be in one or many clusters depending on the method used.
- The disadvantage of hierarchical cluster analysis is that we see only part of the picture. Concepts are placed in one “overall best fit” cluster where it really is in one or many clusters depending on the method used. Hierarchical cluster analysis that many concepts were grouped as one cluster when in reality, they can be in one or many clusters depending on the method used.

Nonhierarchical Clustering

- ORESME allows the researcher to input or “activate” the neuron representing the concept of concepts. Multiple cycles allow for “brainstorming” 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.

Conclusions

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 we try to analyze language and human communication with bounded linear methods.

This study, along with regard to a single commonly experienced event produced some provable results. We used the stem of each year’s output to calculate weighted values. The words which represent the number of words assigned a value to the weighted crime concept. The total signal received by any node is the sum of all the other nodes or

\[ \text{anet} = \sum \text{w}_{ij} \]

References