TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)

TREE REPRESENTATIONS OF ASSOCIATIVE STRUCTURES IN SEMANTIC AND EPISODIC MEMORY RESEARCH

Hervé Abdi
Laboratoire de Psychologie, Dijon

Jean-Pierre Barthélémy
E.N.S.T., Paris

Xuan Luong
Laboratoire de Mathématiques, Besançon

We expose some research in the area of psychology of memory involving proximity or distance matrices. We propose some ways of building up such matrices. Then we detail an algorithm allowing the representation of proximity matrices by an additive tree, and contrast this new algorithm with previous ones. Finally, we examine some results obtained with this method.

1. INTRODUCTION

The general purpose of this paper is to emphasize the utility and describe the use of additive trees in order to describe data collected in the field of the psychology of memory. This paper is threefold: we first describe some research leading to the construction of distance or proximity matrices; secondly we expose and detail the construction of an additive tree as a representation of the original matrix; finally, we examine the results obtained.

The utilization of clustering methods for attesting the organization of memory or revealing its structure has been strongly advocated recently by some authors in different areas of cognitive psychology (see, among others: Miller (1969), Henley (1969), Friendly (1978), Rosenberg et al (1968), (1972), (1982)). Most of the used methods amount to represent the original matrix by an Ultrametric Tree. Recently, there has been an attempt to build some methods leading to representations less stringent than the classical Ultrametric Tree, i.e. the Additive Tree (see Carroll & Chang (1973), Cunningham (1974), (1978); Sattath & Tversky (1977)). We propose hereafter an (economical) heuristic giving an Additive Tree from a proximity matrix and illustrate it with some examples borrowing from our current research or from classical papers.

2. A BUNCH OF EXAMPLES

2.1. BARTLETT & "THE WAR OF THE GHOSTS"

In 1932, Bartlett asked a few subjects to read an American Indian folk tale (named "The War of the Ghosts") and to recall the story on several occasions (a method called "repeated reproduction"); in a variant of the method (i.e. "serial reproduction") a chain of different subjects is used, the first being shown the original text and then recalling for the second subject who would pass it to a third and so on. These classical experiments of Bartlett will serve here to illustrate a set of informatic procedures, the aim of which is to build some distances between texts.

A) Informatic procedures

For reasons of compatibility, the programs are written in Standard Microsoft Basic (Under CP/M), and at least 64 K-Bytes of RAM and a disk Unit are needed. Although these programs include some various possibilities, we will restrict ourself to the part dealing specifically with the construction of metrics between texts.

It must be clear that when we speak of the text given by a subject, we could speak as well of a set of themes or ideas given by a subject providing an adequate coding of the raw data.

The texts are first transformed in a disk file, then for each text we build the Lexicon associated with it. This Lexicon could be either a Boolean Lexicon (i.e. it merely indicates the Presence or the Absence of the item of Vocabulary) or an integer Lexicon (i.e. it indicates the number of Occurrences of each item). From the different Lexicons (Boolean or Integer) we build - by union - a General Lexicon that defines the Vocabulary shared by the different texts.

B) Construction of distances between texts

Depending on the point of view adopted, we could define different distances; as an illustration we examine three ways:

(i) the texts as subsets of the Vocabulary
(ii) the texts as Bi-partitions of the Vocabulary
(iii) a "probabilistic" generalization.

Denote by Li the Lexicon associated with a text Ti, the general Lexicon by $V = \cup_i Li$, and by $\overline{Li}$ the complement of Li in V.

(i) Each (Boolean) Lexicon* is a subset of the Vocabulary and we could use, for example, the well-known distance between sets, the so called cardinal of the symmetric difference:

$$d(Ti,Tj) = |Li \Delta Lj| = |Li \cap \overline{Lj}| + |\overline{Li} \cap Lj|.$$

(ii) $\{Li,\overline{Li}\}$ defines a Bi-Partition of V (i.e. a Partition with two classes), and so does $\{Lj,\overline{Lj}\}$. So, we could use some distances between Partitions (cf. Arabie & Boorman (1973)) or Bi-Partitions, e.g. the distance of the symmetric difference between Bi-partitions that can be expressed as:

$$d(Ti,Tj) = 2(|Li \cap Lj| + |\overline{Li} \cap \overline{Lj}|)(|Li \cap \overline{Lj}| + |\overline{Li} \cap Lj|)$$
$$= 2(|\overline{Li} \Delta \overline{Lj}|)(|Li \Delta Lj|)$$

(iii) In order to take explicitly account of the Integer Lexicons we could look for an extension of (i). With each Ti is associated a probability measure on V (i.e. the frequency of the different words); denote the probability of item x of text Ti by Pi(x); then we find a family of distances by

$$d_q(Ti,Tj) = \sum_{x \in V} [Pi(x)-Pj(x)]^q$$

2.2. FEATURES OF PERSONALITY

This research lays on the border between the work on the organization of the semantic memory and the work upon the "implicit psychology". The purpose is to describe the subjective organization of the qualifiers of the character. As a matter of fact, it has often been noted that we tend to group subjectively some features of character as if we has an "Implicit Theory of Personality" (cf. e.g., Rosenberg et al (1972), Wegner and Vallacher (1977)). In this experiment we select fifty eight qualifiers of the character (using some Thesauruses and a bit of literature,...). These qualifiers are then printed on separate cards and given individually to twenty-eight subjects with the request that he or she sort the cards into piles with the constraint that "the cards in a same pile give the feeling to go together"; subjects were free to choose the number of piles for sorting (for a review of the pro and contra of the sorting method, see Rosenberg (1982)).

---

*Notice that one could " fuzzy " the "boolean" distance in (i) and (ii) by taking the fuzzy equivalent of the union and intersection, i.e. Min and Max.