Revista Brasileira de Física, Vol. 9, Nº 3, 1979

Aplicação da Análise de Agrupamentos Hierárquicos ao Mapeamento Cognitivo de Conceitos Físicos*+

CARLOS ALBERTO DOS SANTOS
Departamento de Física Teórica e Experimental, UFRN, 59000 Natal, RN

e

MARCO ANTONIO MOREIRA
Instituto de Física, UFRGS, 90000 Porto Alegre, RS

Recebido em 25 de Maio de 1979

A "cognitive mapping" was obtained through the application of Hierarchical Clustering Analysis (HCA) to similarity, or proximity, measures provided by concept association tests in the area of Thermodynamics at introductory college level. The basic assumption is that these measures reflect some aspects of the conceptual organization existing in the student's cognitive structure for a given set of concepts. Unlike Multidimensional Scaling which supposes this organization can be represented by an Euclidean configuration, the Hierarchical Clustering Analysis supposes only the existence of a hierarchical organization in the similarity measures.

É apresentado o "mapeamento cognitivo" obtido através de aplicação da Análise de Agrupamentos Hierárquicos (AAH) às medidas de similaridade, ou proximidade, fornecidas por testes de associação numérica e escrita de conceitos na área de Termodinâmica ao nível de Física

* Trabalho parcialmente financiado pela FINEP.

† Versão preliminar e reduzida deste trabalho foi apresentada no IV Simpósio Nacional de Ensino de Física, Rio, 8 a 12 de janeiro de 1979, sob a forma de comunicação intitulada "Mapeamento Cognitivo": ANÁLISE DE AGRUPAMENTOS HIERÁRQUICOS.

Geral. A suposição é a de que tais medidas reflitam aspectos da organização conceitual que o aluno dá a esses conceitos em sua estrutura cognitiva. Diferentemente da Análise Multidimensional onde se supõe que essa organização possa ser representada por uma configuração euclidiana, a Análise de Agrupamentos Hierárquicos supõe apenas a existência de uma organização hierárquica nas medidas de similaridade.

1. INTRODUÇÃO

Santos e Moreira |SM79a| apresentaram dois testes de associação, Teste de Associação Numérica de Conceitos (TANC) e Teste de Associação Escrita de Conceitos (TAEC), que fornecem medidas de similaridade entre conceitos de um determinado conteúdo. No primeiro desses testes, tais medidas são obtidas diretamente: para um dado conjunto de conceitos são formados todos os pares possíveis, os quais são, então, listados em ordem aleatória ao lado de uma escala numérica na qual deve ser assinalado um número que reflita o grau de relacionamento entre os conceitos de cada par segundo a percepção de quem responde o teste. No segundo, as medidas de similaridade são obtidas indiretamente: a cada conceito de um dado conjunto são associados tantos outros conceitos ou palavras quanto possível num certo tempo; a partir das listas de palavras associadas a cada conceito do conjunto, determina-se um coeficiente de relacionamento para cada par de conceitos possível. Em ambos os casos, então, obtém-se um conjunto de medidas de similaridade.

O problema é determinar a estrutura, se houver, inerente ao conjunto de medidas de similaridade obtida através desses testes.

Em outro trabalho Santos e Moreira |SM79b| apresentam uma técnica estatística, Análise Multidimensional (AMD), capaz de ajustar uma configuração em um espaço euclidiano às medidas de similaridade obtidas através de testes de associação de conceitos.

Veremos agora uma técnica alternativa que não impõe a restrição de ajuste em um espaço euclidiano. Enquanto a AMD supõe que os conceitos sejam representados por pontos em um espaço euclidiano, a Análise de Agrupamentos Hierárquicos (AAH) |J67|, supõe apenas a existência de uma estrutura com uma métrica particular, não necessariamente em um espaço físico, concreto.

2. ALGORITMO PARA A AAH

A figura 1 mostra um resultado típico de uma AAH.

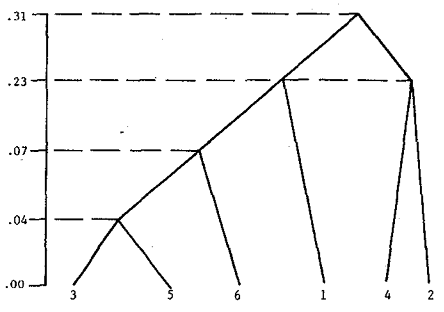

Fig.1 - Resultado típico de uma AAH para seis "conceitos".

Vejamos algumas características desse resultado. No primeiro nível de agrupamento (.00) cada "conceito" é um grupo, de modo que com 6 conceitos sempre teremos 6 grupos distintos no nível .00. A seguir vem o nível .04 com cinco grupos; o par {3,5} é um deles e os outros quatro conceitos são eles próprios os grupos. No nível .07 temos quatro grupos: {3,5,6}, {1}, {4} e {2}. Em .23 temos {3,5,6,1} e {4,2} e, finalmente, no nível .31 temos apenas um grupo: {3,5,6,1,4,2}.

Podemos interpretar esses níveis como distâncias entre os grupos, sendo essas proporcionais aos graus de relacionamento entre os conceitos ou grupos de conceitos.

Nesse sentido, os conceitos que primeiro se agruparem são os mais relacionados entre si. No exemplo dado, os "conceitos" 3 e 5.

No nível .00 sempre teremos tantos grupos quantos forem os conceitos. Esse número vai diminuindo até o valor unitário.

Se num dado agrupamento hierárquico (AH) a formação de grupos for tal que cada conceito se agrega a um grupo já formado, teremos um AH do tipo "linear". O exemplo da figura 1 é "não linear", pois os conceitos 2 e 4 formam um grupo isolado no nível .23.

Vejamos agora como esse esquema de agrupamento hierárquico fornece um tipo particular de distância entre os conceitos, satisfazendo as propriedades de uma métrica.

Sejam n conceitos e uma sequência de $m+1$ agrupamentos, C_0, $C_1, C_2, \ldots, C_m$. A cada agrupamento C_j está associado um nível α_j. C_j é o número de grupos no nível α_j ($\alpha_0 = 0$ e $C_0 = n$).

Supondo que seja conhecido o resultado da figura 1, o problema é determinar a métrica da estrutura, ou, as distâncias entre todos os conceitos.

Dados dois conceitos x e y, sabemos que em C_m eles estão no mesmo grupo. Seja j o menor inteiro do conjunto $[0,1,\ldots,m]$, tal que no agrupamento C_j, x e y estejam no mesmo grupo.

Podemos definir a distância entre x e y, $d(x,y)$, como:

$$d(x,y) = \alpha_j \qquad (1)$$

Na figura 1 temos $d(3,5) = .04$, $d(1,4) = .31$, e assim por diante. A matriz distância para os seis "conceitos" da figura 1 é dada na tabela 1.

Tabela 1 - Matriz distância correspondente à Figura 1.

	1	2	3	4	5	6
1	0	.31	.23	.31	.23	.23
2	.31	0	.31	.23	.31	.31
3	.23	.31	0	.31	.04	.07
4	.31	.23	.31	0	.31	.31
5	.23	.31	.04	.31	0	.07
6	.23	.31	.07	.31	.07	0

Da definição 1 resulta que

$$d(x,x) = \alpha_0 = 0$$
$$d(x,y) = \alpha_j \neq 0 \quad \text{se} \quad x \neq y$$
$$d(x,y) = d(y,x)$$

Para que a função distância $d(x,y)$ satisfaça as proprieda-des de uma métrica, falta apenas mostrar a desigualdade triangular.

Sejam x, y e z três conceitos quaisquer, e sejam

$$d(x,y) = \alpha_j \ ,$$

e

$$d(y,z) = \alpha_k \ .$$

Então, x e y estão no mesmo grupo em C_j e y e z em C_k. Como os agrupamentos são hierárquicos, o grupo que corresponde ao maior dentre j e k inclui o outro.

Seja $\ell = \max[j,k]$. Então, em C_ℓ, x, y e z estão todos no mesmo grupo. Da definição de d, podemos escrever

$$d(x,z) \leq \alpha_\ell$$

e

$$\alpha_\ell = \max[\alpha_j, \alpha_k] \ ;$$

finalmente,

$$d(x,z) \leq \max[d(x,y)\ ,\ d(y,z)] \ . \qquad (2)$$

Por outro lado,

$$\max[d(x,y)\ ,\ d(y,z)] \leq d(x,y) + d(y,z) \ ;$$

então

$$d(x,z) \leq d(x,y) + d(y,z) \ . \qquad (3)$$

A relação (2) é a desigualdade ultramétrica, que é muito mais "forte" que a desigualdade triangular (3). Mostramos, portanto, que dada uma sequência de agrupamentos podemos obter uma função distância que satisfaz as propriedades métricas.

Façamos agora o contrário. Dada uma matriz distância, tal como a da tabela 1, representando alguma métrica d que satisfaz a relação (2), devemos determinar uma sequência de agrupamentos tal como a da figura 1.

No nível .00 temos o agrupamento de referência, com seis grupos ("clusters"), cada "conceito" é um grupo.

O menor elemento da matriz distância é .04, a distância entre os "conceitos" 3 e 5. Isto significa que no nível .04 temos cinco aglomerados: $\{3,5\}$, $\{1\}$, $\{2\}$, $\{4\}$, $\{6\}$.

Na tabela 1, no entanto, 3 e 5 estão exatamente à mesma distância de todos os outros "conceitos". Isto é,

$$d(3,x) = d(5,x), \quad \text{para} \quad x = 1,2,4,6 .$$

O aglomerado $\{3,5\}$ passa, portanto, a ser um novo "conceito", e

$$d(\{3,5\},x) = d(3,x) = d(5,x)$$

Dessa forma, podemos construir nova matriz, e assim por diante, até chegarmos ao último agrupamento, onde todos os conceitos formam um único aglomerado.

Evidentemente o algoritmo desenvolvido até aqui trata de um caso particular, no qual as distâncias satisfazem a desigualdade ultramétrica (relação 2), de modo que

$$d(\{x,y\},z) = d(x,z) = d(y,z) . \tag{4}$$

Em geral, as medidas de similaridade não satisfazem a relação (2), e (4) deve ser escrita como

$$d(\{x,y\},z) = f(d(x,z),\ d(y,z)) . \tag{5}$$

Uma função muito boa, quando se deseja um algoritmo não métrico (que preserva a ordem), é a função "mínimo". Assim, (5) fica

$$d(\{x,y\},z) = \min[d(x,z),\ d(y,z)] \ . \qquad (6)$$

Portanto, dada uma matriz de similaridades, podemos construir uma sequência de agrupamentos hierárquicos, onde as distâncias são calculadas pela relação (6).

Por exemplo, seja o resultado do TANC, após a aprendizagem, apresentado na tabela 2, obtido num experimento feito por Santos |S78| na área de Termodinâmica; pela ordem, temperatura, entropia, calor, energia interna, calor específico, variável de estado, trabalho e equilíbrio térmico.

No nível de referência .00, todos os conceitos estão separados (ver figuras 2 e 3).

O menor valor de matriz na tabela 2 é 1.07, entre Q e U logo, no nível 1.07 temos um grupo formado pelos conceitos Q e U.

Tabela 2 - Matriz distância obtida em um experimento em Termodinâmica.

	T	S	Q	U	C	VE	W	ET
T		1.90	1.93	2.20	2.73	4.20	3.60	1.20
S			3.00	3.13	5.41	1.67	3.57	2.63
Q				1.07	2.70	4.66	2.10	2.21
U					3.80	3.40	1.63	2.71
C						5.83	5.27	4.10
VE							4.47	4.00
W								3.87
ET								

De acordo com o algoritmo determina-se uma matriz 7×7, onde o par Q,U passa a ser um novo "conceito", e as "distâncias" entre todos os outros conceitos permanecem as mesmas da tabela 1.

Assim por exemplo,

$$d[\{Q,U\},T] = \min[d(Q,T),\ d(U,T)] = d(Q,T) = 1.93$$

$$d[\{Q,U\},S] = \min[d(Q,S),\ d(U,S)] = d(Q,S) = 3.00$$

A continuidade do algoritmo resulta na figura 3.

3. EXEMPLOS

As figuras 2e 3 apresentam exemplos de agrupamentos hierárquicos antes e após a aprendizagem, respectivamente, para o TANC no experimento realizado por Santos |S78|.

Observa-se uma modificação relevante no agrupamento após a instrução. A figura 1 mostra um agrupamento "linear", o que caracteriza a inexistência de grupos distintos. Isto é, a organização conceitual na

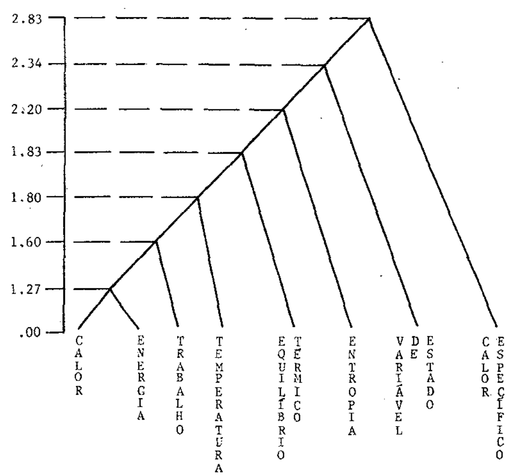

Fig.2 - Resultado da aplicação da AAH ao TANC antes da instrução.

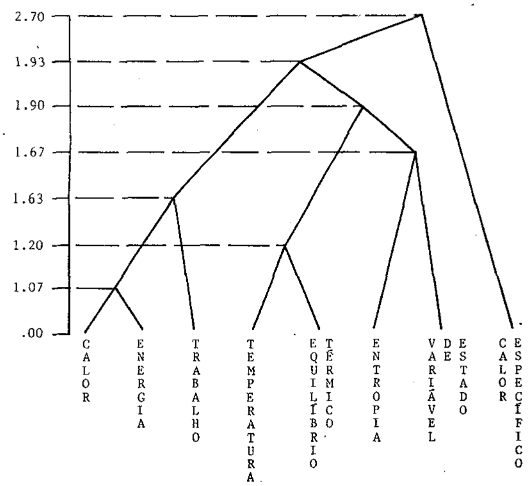

Fig.3 - Resultado da aplicação da AAH ao TANC após a instrução.

estrutura cognitiva apresenta uma distribuição aproximadamente homogênea.

A figura 2, ao contrário, apresenta uma organização hierárquica, com o formato de grupos distintos, particularmente, os grupos

{Q,U} no nível 1.07, {T,ET} em 1.20,

{Q,U,V} em 1.63, e {T,ET,S,VE} em 1.90.

Os resultados relativos ao TAEC, antes e após a instrução, figuras 4 e 5, sugerem a existência de uma componente semântica no referido teste |S78|. A influência dessa componente ocorre, quando dois conceitos têm denominações que se assemelham, na linguagem comum na experiência cotidiana. Desse modo, além das relações que dois conceitos possam ter no contexto da Física, eles podem ser associados por lembrarem experiências semelhantes na vida diária.

É possível, por exemplo, que o grupo {T,Q}, da figura 4 tenha se formado em consequência de tal componente.

Da mesma forma, na estrutura da figura 5, talvez o aparecimento do grupo {T,ET} em primeiro lugar sugira também a influência da componente semântica.

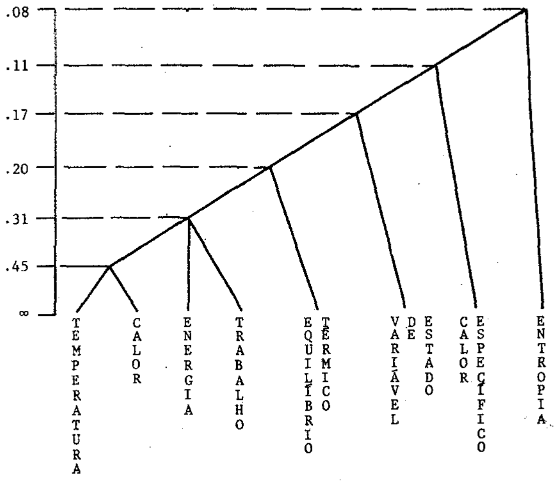

Fig.4 - Resultado da aplicação da AAH ao TAEC antes da instrução.

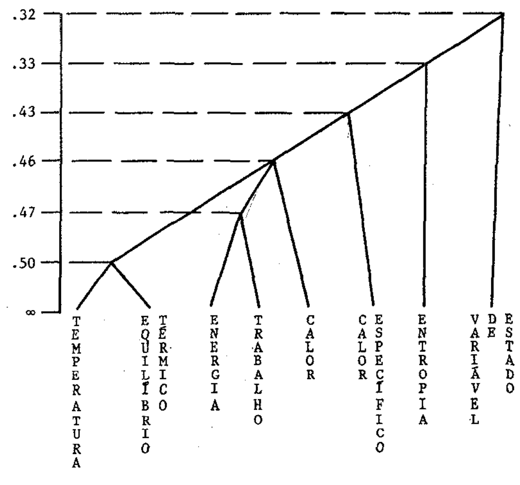

Fig.5 - Resultado da aplicação da AAH ao TAEC após a instrução.

4. CONCLUSÃO

Considerando os exemplos apresentados e vários outros resultados |S78|, podemos dizer que os testes usados convergem razoavelmente bem, e que nos parece conveniente um estudo relativo à influência semântica no TAEC e suas implicações na estrutura cognitiva.

Além disso, parece-nos também que a AAH é uma técnica potencialmente útil quando se está interessado em informações qualitativas acerca da aprendizagem de conceitos. Por exemplo, se usados como diagnóstico os "mapas" obtidos com aplicação da AAH aos resultados dos testes de associação podem detectar falhas no entendimento de relações entre conceitos, as quais, por sua vez, podem influir no desempenho do aluno.

REFERÊNCIAS

J67 Johnson, S.C. Hierarchical clustering schemes. Psychometrika, Chicago, *32*(3):241-54, Sept. 1977.

S78 Santos, C.A. *Aplicação da Análise Multidimensional e da Análise de Agrupamentos Hierárquicos ao Mapeamento Cognitivo de Conceitos Físicos*. Dissertação de Mestrado, Instituto de Física, UFRGS, 1978.

SM79a Santos, C.A. & Moreira, M.A. *Mapeamento Cognitivo: INSTRUMENTOS DE MEDIDA*. Comunicação apresentada no IV Simpósio Nacional de Ensino de Física, Rio de Janeiro, 8 a 12 de janeiro de 1979.

SM79b Santos, C.A. & Moreira, M.A. *Mapeamento Cognitivo: ANÁLISE MULTIDIMENSIONAL*. Comunicação apresentada no IV Simpósio Nacional de Ensino de Física, Rio de Janeiro, 8 a 12 de janeiro de 1979.