

MÉTODOS DE CONSENSO EM ANÁLISE DE DADOS

Anabela Cardoso Marques

Escola Superior de Tecnologia do Barreiro
Instituto Politécnico de Setúbal

RESUMO

Neste trabalho apresenta-se alguns métodos de consenso que permitem ao investigador encontrar uma classificação hierárquica única para um conjunto de r classificações hierárquicas, ou r dendrogramas, ou r árvores.

Palavras-chave: Métodos de Consenso; Análise Classificatória; Classificação Hierárquica

1. INTRODUÇÃO

Geralmente um consenso está associado a uma tomada de decisão, a qual é entendida como sendo o resultado de um acordo em termos de opinião ou testemunho. Assim um consenso pode ser obtido de três formas diferentes: por maioria, por unanimidade ou por ponderação de votos, onde se inclui um consenso médio.

2. CRITÉRIOS PARA A OBTENÇÃO DE UM CONSENSO

De acordo com Black (1963), o critério básico para a obtenção de um consenso pode ser especificado da seguinte forma:

Suponhamos que temos um conjunto de k elementos, $E = \{1, \dots, k\}$, aos quais é apresentada uma questão e sobre a qual se tem de expressar uma opinião, tendo-se m alternativas de resposta, $X = \{x_1, \dots, x_m\}$. A decisão tomada para cada elemento pode ser especificada num perfil $P = (p_1, \dots, p_k)$, onde p_i é a alternativa escolhida pelo i -ésimo elemento, $1 \leq i \leq k$. Para cada alternativa x_j , $n_j(P)$ conta o nº de ocorrências de x_j em P .

O critério para obter um consenso por unanimidade pode ser representado pela função U para cada perfil P , da seguinte forma:

$$U(P) = \begin{cases} x_j & \text{se } n_j(P) = K \\ \emptyset & \text{caso contrário} \end{cases}$$

No entanto, se optarmos por uma tomada de decisão por maioria, basta considerar para cada perfil P , a função MAJ, tal que:

$$MAJ(P) = \begin{cases} x_j & \text{se } n_j(P) > K/2 \\ \emptyset & \text{caso contrário} \end{cases}$$

No entanto as duas definições anteriores não permitem encontrar um consenso para o caso em que temos o mesmo número de respostas para as alternativas escolhidas com maior frequência.

Para tentar dar resposta a estas situações, surge uma outra definição de consenso, que tem por base a escolha de um “conjunto” médio, fazendo com que a tomada de decisão seja constituída por uma ou por várias opções, desde que estas se encontrem empatadas.

Seja d uma distância definida em X^2 , onde X é o conjunto das alternativas possíveis para a tomada de decisão sobre um determinado acontecimento.

O critério para a obtenção dum consenso utilizando uma regra média é dado pela função MED para um perfil P , cujo resultado é obtido através de:

$$\text{MED}(P) = \{x \in X : \sum_{p \in X} d(x, p) \text{ é mínimo}\}$$

onde d é tal que:

$$d(x, y) = \begin{cases} 0 & \text{se } x = y \\ 1 & \text{se } x \neq y \end{cases}$$

2. EVOLUÇÃO E EXTENSÃO DO CONSENSO

Com o passar das décadas foram feitas diversas adaptações e melhoramentos ao conceito inicial de consenso, de forma a se poder estender a sua aplicação a diferentes áreas da matemática nomeadamente à análise de dados.

Por exemplo, Barthelemy e Monjardet (1981), mostraram como o consenso médio pode ser aplicado a problemas que envolvem relações: completas, de ordem, de simetria, de equivalência,...

Barthelemy, Leclerc e Monjardet (1986), deram uma visão compreensiva de como um conjunto ordenado é usado na obtenção de uma classificação de consenso, muitos dos seus resultados são aplicáveis a problemas de consenso que envolvem partições, dendrogramas, árvores, árvores ordenadas, árvores genéticas.

A árvore é a representação gráfica utilizada pelos diversos métodos para a obtenção da classificação hierárquica de consenso, em Adams (1972), podemos ver um método computacional que converte a informação contida em vários perfis, numa árvore de consenso, sendo esta constituída pela informação que é partilhada/idêntica em todos esses perfis.

Consideremos a seguinte notação:

Seja $N = \{1, \dots, n\}$ o conjunto dos elementos a classificar cuja informação é registada nas n etiquetas das folhas de uma árvore. Dizemos que temos uma árvore T de N se tivermos uma família de subconjuntos de N tais que:

- $N \in T$,
- $\emptyset \notin T$,
- $\{x\} \in T$ para todo $x \in N$

Utilizando esta notação, Sokal e Rohlf, em 1981, propuseram que para cada perfil $P=(T_1, \dots, T_k)$, de k árvores com n elementos, se obtivesse, uma outra árvore $U(P)$ de n elementos, tal que:

$$U(P)=T_1 \cap T_2 \cap \dots \cap T_k$$

Esta nova árvore é designada por árvore de consenso estrita, uma vez que a sua obtenção provém da recolha da informação que é comum às k árvores, estando desta forma associada à obtenção de um consenso por unanimidade.

Em 1981, Margush e McMorris, introduziram a seguinte regra para a obtenção de uma árvore de consenso por maioria - MAJ(P), onde:

$$MAJ(P)=\{c \subseteq N: c \in T_j \text{ para mais de metade das árvores de } P\}$$

Em 1983 McMorris e Neumann, caracterizaram axiomáticamente a família (M_L) dos métodos de consenso para um conjunto de árvores com n elementos a classificar, aparecendo as metodologias para a classificação de consenso apenas por volta de 1986, em revistas como "Journal of Classification".

No final da década de 80, Barthelemy and McMorris, caracterizaram axiomáticamente o método de consenso médio para um conjunto de árvores com n elementos a classificar, a partir de uma medida da dissemelhança (d), apresentando a seguinte formulação:

$$Med(P) = \{T: T \text{ é uma árvore, } \sum_{p \in X} d(T, p) \text{ é mínima}\}$$

Onde para um conjunto de árvores (P) , a função $Med(P)$ encontra a árvore que comparada com todas as outras fornece um valor mínimo para uma determinada medida de dissemelhança (d) escolhida pelo investigador.

3. CONSENSO EM ANÁLISE CLASSIFICATÓRIA

Em Gordon (1999), podemos ver que dadas t ($t \geq 2$) classificações hierárquicas $\{T_r: T_r \text{ é uma árvore } (r=1, \dots, t)\}$ sobre o mesmo conjunto de n unidades estatísticas a classificar, torna-se vantajoso obter uma árvore de consenso única que sintetize a informação contida nas t classificações originais. Esta árvore de consenso tem de ser tal que a informação contida na sua estrutura deve fornecer ao investigador um resumo fiável das interligações existentes entre essas unidades estatísticas.

As t classificações originais podem ter sido obtidas:

- por aplicação de t procedimentos de análise classificatória ao conjunto das matrizes de distâncias entre as unidades estatísticas;
- a partir do resultado da análise de t investigadores diferentes;
- a partir da análise de um único investigador, mas considerando diferentes conjuntos de variáveis na descrição das unidades estatísticas.

A metodologia das árvores de consenso está relacionada com a síntese de árvores a qual tem em conta as classes consistentes de cada árvore.

A árvore de consenso estrita, definida por Sokal e Rohlf, apresenta as classes que aparecem em todas as t árvores originais, o que por vezes faz com que esta árvore seja constituída por um número reduzido de classes.

4. CONCLUSÕES

As técnicas dos métodos de consenso podem ser vistas de uma forma simples e abstracta como um utensílio para a análise de dados multidimensionais. O investigador, perante um conjunto de t classificações hierárquicas e segundo um determinado critério (unanimidade; maioria; médio) pode obter uma única classificação hierárquica que resuma essas t classificações originais.

A escolha do critério a usar na obtenção do consenso depende do grau de precisão que o investigador pretende alcançar.

5. REFERÊNCIAS

ADAMS, E. N., III (1972) – “Consensus techniques and the comparison of taxonomic trees”. *Systematic Zoology* 21, 390-397.

BARTHELEMY, J.P., LECLERC, B., and MONJARDET, B. (1986) – “On the use of ordered sets in problems of comparison and consensus of classifications”. *J. of Classification* 3, 187-224.

BARTHELEMY, J.P., and McMORRIS, F.R. (1986) – “The median procedures for n -trees”. *J. of Classification* 3, 329-334

BARTHELEMY, J.P., and MONJARDET, B. (1981)– “The median procedure in cluster analysis and social choice theory”. *Mathematical Social Sciences* 1, 235-267.

BLACK, D. (1963) – “The theory of committees and Elections”, *University Press, Cambridge, England*

MARGUSH, T., and McMORRIS, F. R. (1981) – “Consensus n -trees”. *Bulletin of Mathematical Biology* 43, 239-244.

MCMORRIS, F. R., and NEUMANN, D. (1983)– “Consensus functions defined on trees”. *Mathematical Social Sciences* 4, 131-136.

SOKAL, R. R., and ROHLF, F.J. (1981) – “Taxonomic congruence in the Leptopodomorpha re-examined”. *Systematic Zoology* 30, 309-325.