4 TEORIA MATEMÁTICA DA COMUNICAÇÃO DE SHANNON

A Teoria Matemática da Comunicação, ou Teoria da Comunicação de Shannon foi proposta por Claude Shannon no final da década de 1940 como forma de sistematizar o conhecimento necessário ao entendimento da eficiência em sistemas de comunicação (KHINCHIN, 1957, p. 30). A TMC não foi considerada como adequada para aplicações em cartografia por alguns autores como Salichtchev (1973)[1], Robinson e Petchenik (1976)[2] citados por Bjorke (BJORKE, 1996, p.78), em função de que no meio de comunicação, objeto da Cartografia, não ocorre perda de informação, mas pelo contrário, nestes sistemas durante sua leitura e observação são suscitadas impressões e mesmo relacionamentos que não foram de forma consciente projetados pelo cartógrafo. No entanto, Bjorke (BJORKE, 1996, p.79), apresenta o conceito de entropia, da TMC, como apropriado para aplicações no nível sintático da comunicação cartográfica.

Um sistema de comunicação é composto por um gerador de informação, um meio de transmissão e um receptor (Fig. 4.1). O gerador de informação é caracterizado como um conjunto X de n eventos aos quais se podem associar valores de probabilidade $P(X)$. A recepção também é caracterizada como um conjunto de mesmo número n de eventos aos quais estão associados valores de probabilidades $P(Y)$. Neste sistema $P(X)=\{p_{x1}, p_{x2}, ..., p_{xn}\}$ representa o conjunto de valores de probabilidades associadas aos elementos do conjunto X de informação original, e $P(Y)=\{p_{y1},p_{y2},...,p_{yn}\}$ representa o conjunto de valores de probabilidade associados ao conjunto Y de informação recebida.

Durante o processo de transmissão pode ocorrer perda de informação. Esta perda de informação se reflete nos valores de probabilidade associadas aos elementos do conjunto Y. A perda de informação é mostrada por meio de dois escalares denominados equivocação $H(Y,X)$ e informação útil $R(X)$. Quanto maior for a

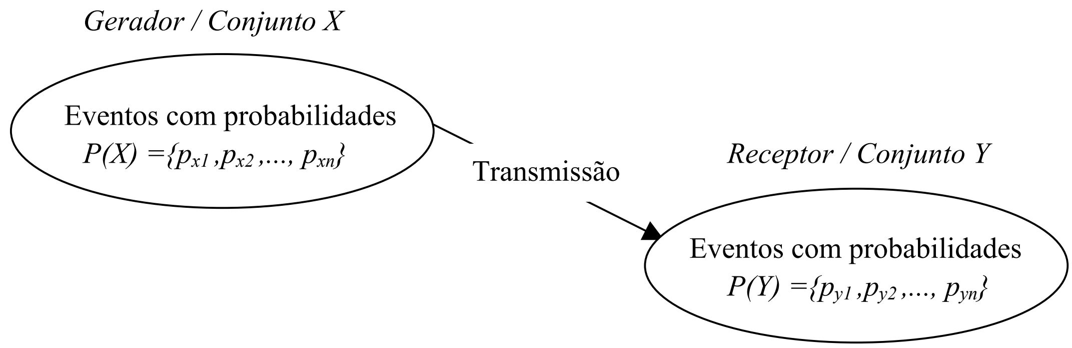
---

[1] SALICHTCHEV, K.A. 1973, Some reflections on the subject and method of cartography after the Sixth International Cartographic Conference. *Canadian Cartographer* 10:106-11.

[2] ROBINSON, A.H. , PETCHENIK, B.B. 1976. *The nature of maps*. Essays toward understanding maps and mapping. Chicago: The University of Chicago Press.

equivocação menor será a informação útil. Não ocorre perda de informação num sistema de comunicação quando o valor de probabilidade de cada elemento do conjunto *Y* for igual a cada elemento do conjunto *X* . De modo geral sempre ocorre perda de informação, e com isso as probabilidades dos elementos do conjunto *Y* são diferentes das probabilidades dos elementos do conjunto *X*.

FIGURA 4.1 - SISTEMA DE COMUNICAÇÃO DO PONTO DE VISTA DA TMC



Pela TMC, a cada evento do conjunto *X* deve corresponder uma probabilidade p_x associada. Isto quer dizer que a aplicação da TMC requer a identificação dos dados que compõem o conjunto *X*, e requer a determinação das probabilidades p_{xi} a eles associadas. Para um conjunto de eventos as probabilidades existem no intervalo [0,1], e a soma de todas as probabilidades de um conjunto de eventos deve ser igual a 1. A determinação das probabilidades dos eventos do conjunto *X* deve ser feita com base em alguma propriedade dos dados do conjunto. Por exemplo, as probabilidades para um conjunto de *n* pontos com coordenadas planimétricas pode ser dada por $p_{xi} = 1 / n$, isto é, neste caso todos os pontos tem a mesma probabilidade.

A determinação das probabilidades dos elementos do conjunto *X* pode ser feita mediante a aplicação de alguma função que relacione os dados segundo algum critério ou propriedade. Por exemplo, a um conjunto de *n* pontos pode ser atribuída a mesma probabilidade *1/n* a cada elemento do conjunto. No entanto, se for detectada e existência de alguma propriedade que diferencie os pontos dentro do conjunto, podem ser atribuídos diferentes valores de probabilidades.

A entropia $H(X)$, é um escalar associado a um conjunto X de n eventos, representados por meio de suas probabilidades $P(X)=\{p_{x1}, p_{x2}, ..., p_{xn}\}$, dado pela expressão:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \ \log_2 \ p(x_i) \tag{4.1}$$

O comportamento dos valores de probabilidades do conjunto de eventos determina as diferentes situações de entropia. A entropia $H(X)$ de um conjunto de eventos tem as seguintes propriedades:

a) um conjunto de n eventos em que $n-1$ eventos têm probabilidade nula e apenas um evento tem probabilidade igual a 1, $P(X)=\{0,0,..,0,1\}$ tem entropia $H(X) = 0$. Isto quer dizer que para este conjunto não há incerteza sobre que evento ocorreu;
b) para um conjunto de n eventos em que todos os eventos têm a mesma probabilidade $p(x_i)$, a entropia será máxima, $H(X) \Rightarrow max$; e
c) para um conjunto de n eventos em que as probabilidades $p(x_i)$ assumem diferentes valores, a entropia será um valor diferente de zero, porém, menor do que o valor máximo, $0 < H(X) < max$.

A determinação da eficiência de um sistema de comunicação requer a formalização de um conceito de similaridade a ser aplicado aos dados. A similaridade $\mu(y,x)$ é um valor definido para o intervalo $[0,1]$ obtido como resultado da avaliação da semelhança entre um elemento do conjunto Y com um elemento do conjunto X. Quanto maior for a possibilidade dos dois elementos se confundir maior será seu grau de similaridade. Existe um elemento do conjunto Y ao qual corresponde um elemento do conjunto X em que ocorre o máximo grau de similaridade, "todo elemento do conjunto Y é similar a si próprio, e o valor do grau de similaridade nesse caso é o valor máximo de $\mu(y,x)$, isto é, o valo *1*".

A visualização da ocorrência de similaridades entre os elementos dos conjunto X e Y pode ser feita usando a figura de uma matriz quadrada que tem dimensão igual ao número *n* de elementos do conjunto *X={k, l, m,n}*, (Fig. 4.2). A diagonal principal da matriz é composta por valores $\mu(y,x)=1$ uma vez que cada elemento do conjunto é similar a si próprio no grau máximo. Um elemento *m* da diagonal principal pode ser similar a um elemento *n* também da diagonal principal em um grau entre *[0,1]*. O valor deste grau de similaridade $\mu(m,n)$ é posicionado tanto na linha *m* e coluna *n*, quanto na linha *n* e coluna *m*, isso quer dizer que a matriz com valores de grau de similaridade é simétrica (Fig. 4.2b ou c).

FIGURA 4.2 - MATRIZES DE GRAUS DE SIMILARIDADE $\mu(Y,X)$ PARA 4 ELEMENTOS (*K*, *L*, *M*, *N*)

$$\begin{pmatrix} \boldsymbol{\mu(k,k)} & & & \\ & \boldsymbol{\mu(l,l)} & & \\ & & \boldsymbol{\mu(m,m)} & \\ & & & \boldsymbol{\mu(n,n)} \end{pmatrix}$$

a) *k, l, m* e *n* similares a si próprios

$$\begin{pmatrix} \boldsymbol{\mu(k,k)} & & & \\ & \boldsymbol{\mu(l,l)} & \mu(l,m) & \\ & \mu(l,m) & \boldsymbol{\mu(m,m)} & \mu(m,n) \\ & & \mu(m,n) & \boldsymbol{\mu(n,n)} \end{pmatrix}$$

b) *l* similar a *m* e *m* similar a *n*

$$\begin{pmatrix} \boldsymbol{\mu(k,k)} & \mu(k,l) & \mu(k,m) & \mu(k,n) \\ \mu(k,l) & \boldsymbol{\mu(l,l)} & \mu(l,m) & \mu(l,n) \\ \mu(k,m) & \mu(l,m) & \boldsymbol{\mu(m,m)} & \mu(m,n) \\ \mu(k,n) & \mu(l,n) & \mu(m,n) & \boldsymbol{\mu(n,n)} \end{pmatrix}$$

c) *k*, *l*, *m* e *n* similares entre si em algum grau no intervalo *[0,1]*

As posições das matrizes em que não existem valores $\mu(y,x)$ representam similaridades nulas. Quando os elementos *k*, *l*, *m* e *n* são similares apenas a si próprios, a matriz de similaridades resulta numa matriz identidade e a entropia do conjunto *Y* será a mesma do conjunto *X* (Fig. 4.2a). Quando alguns elementos do conjunto *Y* podem ser confundidos ou tem algum grau de similaridade com elementos do conjunto *X*, a matriz de similaridades apresenta valores de similaridade diferentes de zero fora da diagonal principal (Fig. 4.2b). Neste caso, as entropias *H(X)* e *H(Y)* são diferentes. Quando todos os elementos do conjunto *Y* podem ser confundidos ou tem algum grau de similaridade com os elementos de X a matriz de similaridades é formada apenas por elementos não nulos (Fig. 4.2c). Também neste caso as entropias $H(X)$ e $H(Y)$ serão diferentes. Alto grau de similaridade resulta em valor próximo de *1* e baixo grau de similaridade resulta num valor próximo de zero.

As probabilidades dos elementos do conjunto Y devem refletir a existência de similaridade determinada com a aplicação da função de similaridade. Os valores de grau de similaridade tem existência no intervalo $[0,1]$. Para cada um dos n elementos do conjunto X podem existir n valores de grau de similaridade. Estes graus de similaridade são obtidos por aplicação de uma função de similaridade a cada par de elementos possível de realizar no conjunto de n elementos. A soma de todos os valores dos graus de similaridade de um dado elemento é maior ou igual a um, $\sum \mu(y,x) \geq 1$. Esta soma somente será igual a *1* se o elemento comparado for similar apenas a si próprio. A probabilidade de transição ou probabilidade condicional é definida como um valor existente no intervalo $[0,1]$, e é calculada usando os valores de similaridade na seguinte expressão:

$$p(y,x) = \frac{\mu(y,x)}{\sum \mu(y,x)} \tag{4.2}$$

Nesta expressão, $p(y,x)$ representa a probabilidade condicional ou probabilidade de transição, $\mu(y,x)$ representa o valor do grau de similaridade que o elemento representado tem com o elemento tomado como base na comparação, e $\sum\mu(y,x)$ representa a soma de todos os graus de similaridade obtidos para um elemento de X tomado como base.

Os dados de probabilidade de transição podem ser organizados em uma matriz que tem as mesmas dimensões da matriz de graus de similaridade. A diferença entre os valores de grau de similaridade e probabilidade de transição é que a soma de uma linha ou de uma coluna na primeira pode resultar num valor maior ou igual a *1*, e a soma de uma linha ou coluna na matriz de probabilidades de transição deve necessariamente resultar em *1* para que esteja de acordo com o conceito de probabilidade.

O conjunto de dados transmitido ou representado (segundo a terminologia cartográfica), tem valores de probabilidades associado. Estes valores de probabilidade são dependentes dos valores de probabilidade *P(X)* associada aos elementos do conjunto original e dependentes da possibilidade de serem confundidos entre si, como expresso pela função de similaridade e da probabilidade de transição *p(x,y)*. A probabilidade dos elementos representados é dada pela expressão matricial:

$$P(Y) = P(X)\ .\ P(Y,X) \tag{4.3}$$

onde, *P(X)* é o vetor que contém os valores das probabilidades dos elementos originais do conjunto *X* e *P(Y,X)* é uma matriz que contém os valores das probabilidades condicionais individuais *p(y,x)*.

Quando a matriz de similaridades é uma matriz identidade, isto é nenhum elemento se confunde com nenhum outro elemento na representação, os valores das probabilidades dos elementos dos conjuntos *X* e *Y* são iguais, isso quer dizer que o processo de transmissão da informação não proporciona perda. Quando o sistema de transmissão introduz ruído ou perda, a matriz de similaridades, e por conseqüência, a matriz de probabilidades condicionais tem elementos não nulos fora da diagonal principal, e os valores das probabilidades *P(Y)* resultam diferentes dos valores das probabilidades *P(X)*.

A equivocação *H(Y,X)* é um escalar que sintetiza o grau de equívocos que ocorre na transmissão de dados em um sistema avaliado por meio de uma função de similaridade. O valor da equivocação é tanto maior quanto maior for a presença de valores de grau de similaridades diferentes de zero na matriz de similaridades. Este valor permite concluir acerca da eficiência do sistema de comunicação. Para o caso em que se tenha matriz de similaridades igual a uma matriz identidade, a equivocação será zero, porque neste caso não existem similaridades entre os elementos do conjunto *Y*. No sentido oposto, quando existem valores de similaridade entre os elementos do conjunto *Y*, então, pode haver equívoco ao ser realizada a sua interpretação. O escalar equivocação é obtido de:

$$H(Y,X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y,x) \log_2 p(y,x) \tag{4.4}$$

Nesta expressão, *p(x)* representa o valor de cada uma das probabilidades dos elementos do conjunto *X*. Os valores de *p(y,x)* representam as probabilidades condicionais dos elementos do conjunto *Y*. A equivocação aumenta com o aumento da ocorrência de similaridades.

A informação útil *R* é o outro escalar que permite inferir acerca da eficiência do sistema de comunicação. A informação útil *R* é diretamente proporcional à eficiência do sistema, ou seja, tem comportamento inverso ao comportamento da equivocação. Quanto maior a equivocação menor a informação útil e vice-versa. O valor da informação útil *R* é dado por:

$$R = H(Y) - H(Y,X) \tag{4.5}$$

Para o caso da matriz de similaridades resultar na matriz identidade o valor da informação útil será máximo. Em qualquer outra situação da matriz de similaridade a informação útil será menor. A informação útil será igual a zero quando todo elemento do conjunto *Y* se confundir com todo elemento do conjunto *X*. Isso resultará numa matriz de similaridades que tem todos os elementos iguais a *1*. Este será o caso mais desfavorável.

A aplicação da TMC a um sistema de comunicação tem por finalidade a determinação da capacidade *C* do canal de comunicação que é dada como o máximo valor de informação útil *R* para o sistema. A utilização de uma função que expresse a similaridade entre os elementos produzidos como resultado da comunicação trazem informação acerca de seu nível de similaridade.

Bjorke (BJORKE 1996, p.87) relaciona a TMC com a Cartografia na fase de projeto cartográfico para a verificação da eficiência alcançada na comunicação quando uma representação é produzida iterativamente. O mesmo autor propõe uma sistematização da aplicação da TMC a problemas de cartografia, em que define quatro passos a serem satisfeitos. Estes passos são: a determinação da fonte de dados; o estabelecimento do modelo estocástico; o estabelecimento do modelo de entropia; e a definição do critério de rejeição.

Na aplicação da TMC a problemas de Cartografia, os aspectos mais importantes são a determinação da função que expressa a similaridade observável ou presente nos dados e a determinação do valor do(s) parâmetro(s) para esta função de similaridade. A função de similaridade representa a avaliação de um conjunto de dados de cartografia segundo algum ponto de vista, expressão de alguma condição ou critério cartográfico para comunicação de dados.