MÔ HÌNH TOÁN HỌC SHANNON CỦA THÔNG TIN VÀ TỶ SỐ THÔNG TIN

Thái Thanh Sơn*, Thái Thanh Tùng*

Email: tttung@hou.edu.vn

Ngày tòa soạn nhận được bài báo: 01/11/2023

Ngày phản biện đánh giá: 17/06/2024

Ngày bài báo được duyệt đăng: 26/06/2024

DOI: 10.59266/houjs.2024.409

Tóm tắt: *Trong nửa thế kỷ gần đây, thuật ngữ Thông tin – Information - đã được sử dụng một cách rất phổ biến mọi lúc, mọi nơi trên toàn thế giới. Từ những nhà khoa học kỹ thuật chuyên ngành cho đến những người dân bình thường, hàng ngày đều nói đến thông tin, quan tâm đến thông tin và luôn sử dụng thông tin trong mọi lĩnh vực của cuộc sống của mình. Tuy nhiên, không thể dễ dàng có căn cứ để đưa ra một "định nghĩa" chính xác cho khái niệm đó. Trong bài báo này trước tiên chúng tôi cố gắng trình bày lại một cách tóm tắt và đầy đủ mô hình toán học Shannon và đưa ra những minh hoạ cụ thể, nhằm hỗ trợ những người bước đầu tìm hiểu về Công nghệ thông tin tiếp cận khái niệm quan trọng hàng đầu trong chuyên ngành của mình. Đặc biệt, tiếp đó đã xem xét khái niệm Tỷ số thông tin giữa hai quan sát – the Information Ratio between two observations – có vai trò hết sức quan trọng khi khảo sát sự liên quan giữa các sự kiện ngẫu nhiên không định lượng – một lĩnh vực nghiên cứu còn chưa được khai thác đầy đủ.*

Từ khóa*: thông tin, độ bất định thông tin, lý thuyết thông tin truyền thông, tỷ số thông tin.*

I. Đặt vấn đề

Vì rằng về bản chất khái niệm "thông tin" là một khái niệm nguyên thuỷ (Primitive notion) trừu tượng, cho nên chỉ có thể mô tả những thuộc tính đặc trưng của khái niệm đó thông qua một hệ tiên đề được công nhận. Điều này hoàn toàn nằm ngoài khả năng tiếp cận của đa số người sử dụng bình thường cho nên điều không đáng ngạc nhiên là khi ta hỏi một đối tượng, kể cả người có thể đã làm việc thực sự khá lâu trong một số chuyên ngành ứng dụng của công nghệ thông tin (CNTT), rằng: -"Thông tin là gì?" – thì rất ít khi nhận được câu trả lời thoả đáng. Đối với các đối tượng không chuyên nghiệp – kể cả trong một vài tài liệu nhập môn cho những môn học như: Nhập môn Tin học, CNTT, kỹ thuật máy tính v..v.. ở cấp học thấp - người ta thường cố gắng đưa ra những cách "mô tả, giải thích"…

* Trường Đại học Mở Hà Nội

nào đó, không thể xem là những định nghĩa chính xác, khoa học, mà có khi còn có thể có sai sót.

Thông thường, người ta có thể "mô tả" một cách trực quan rằng: - Thông tin là toàn bộ những "cái gì đó" mà con người cần "thu thập" để mang lại cho họ những "hiểu biết" về thế giới tự nhiên và xã hội con người đang diễn ra quanh mình.

Năm 1948, Claude Shannon trong bài: - *A mathematical model of Information* - [3] - *Một mô hình toán học của Thông tin* – đã đề xuất một cách tiếp cận chính xác, hợp lý và còn có tính định lượng cho khái niệm Thông tin. Tiếp đó, hàng loạt công trình nghiên cứu chẳng hạn như của Amiel Feinstein trong *Information Theory and its applications* – [4] - Lý thuyết thông tin và các ứng dụng của nó - (1954) - đã tiếp tục phát triển hoàn chỉnh đề xuất của C. Shannon, và đến nay mô hình đó đã trở thành cách tiếp cận chính xác duy nhất được sử dụng trong mọi nghiên cứu lý thuyết và ứng dụng của ngành Công nghệ thông tin.

II. Đặc trưng trong lý thuyết thông tin

Lý thuyết thông tin dựa trên lý thuyết xác suất thống kê, trong đó thông tin định lượng thường được mô tả dưới dạng độ đo thông tin là bit. Lý thuyết thông tin thường liên quan đến các biện pháp phân phối thông tin liên quan đến các biến ngẫu nhiên. Một trong những biện pháp quan trọng nhất được gọi là entropy, tạo thành khối xây dựng của nhiều biện pháp khác. Entropy cho phép định lượng các phép đo thông tin trong một biến ngẫu nhiên duy nhất. Một khái niệm hữu ích khác là thông tin được xác định trên hai biến ngẫu nhiên, mô tả thước đo thông tin chung giữa các biến đó, có thể được sử dụng để mô tả mối tương quan của chúng.

2.1. Độ bất định – Entropy – của một quan sát ngẫu nhiên.

Tuy rằng khái niệm thông tin rất trừu tượng và khó hiểu nhưng trong đời sống thông thường người ta vẫn nói đến chuyện "thu thập thông tin "thậm chí vẫn có thể so sánh, chẳng hạn như: "tờ báo hôm nay nhiều thông tin hơn tờ báo hôm qua…". Trước hết, ta xét một sự kiện ngẫu nhiên A_i nào đó xảy ra trong thiên nhiên hay trong xã hội, chẳng hạn:" - Ngày mai có mưa không? – Gieo con xúc xắc lần này có ra mặt "lục" không? ". Khi sự kiện A_i đó chưa xảy ra (chưa đến ngày mai, chưa gieo xúc xắc) thì rõ ràng ta chưa có *"hiểu biết"* gì về việc A_i có xảy ra hay không cả: Có một sự "*mù mờ, không minh bạch - opacity, non-transparency* - về sự kiện A_i, ta gọi đó là độ bất định – entropy – của A_i, ký hiệu là $H(A_i)$. Một cách trực quan, ta thấy ngay là dộ bất định – tính mù mờ - của A_i *tỷ lệ nghịch với* xác suât xuất hiện của A_i, thật vậy: nếu sự xuất hiện của A_i là hầu như chắc chắn: $P(A_i) = 1$ thì không còn gì là bất định nữa: $H(A_i) = 0$, ngược lại nếu khả năng xuất hiện của A_i rất thấp thì $H(A_i)$ lại rất cao. Để thuân tiện cho một số việc tính toán sau này, Shannon đã "đo" entropy của sự kiện A_i bằng một đại lượng tỷ lệ nghịch vơi p_i hay là tỷ lệ thuận với ${}^1/p_i$ là:

$$H(A_i) = \log_a {}^1/p_i = -\log_a p_i \qquad (1)$$

(trong đó a là một cơ số a > 1 nào đó, sau này sẽ sử dụng logarit cơ số 2: a = 2)

Giả sử S là một quan sát ngẫu nhiên, có n kết cục có thể xảy ra $\{A_1, A_2, \ldots, A_n\}$ với xác suất tương ứng là $p_1, p_2, \ldots, p_n$. Ta goi: *Entropy – độ bất định – của quan sát S* ký hiệu là *H(S)* là *kỳ vọng toán học của cảc entropy của* các sự kiện A_1, A_2, ... A_n:

$$H(S) = \Sigma_i\, p_i\, H(Ai) = \Sigma_i\, p_i \log_a {}^1/p_i \quad (2)$$
$$= - \Sigma_i\, p_i \log_a p_i$$

Rõ ràng nếu trong quan sát ngẫu nhiên S nếu chỉ có 1 sự kiện A_1 có $p_1 = 1$ còn các sự kiện khác đều có pi = 0 thì quan sát S không còn gì là bất định nữa: Trong trường hợp này ta có H(S) = 0. Ngược lại nếu trong quan sát S, các kết cục A_i đều đồng khả năng $p_i = 1/n$ thì H(S) đạt cực đại:

$$\text{Max } H(S) = \Sigma_i\, {}^1/_n \log_a n = \log_a n \quad (3)$$

Chẳng hạn, khi xảy ra một trận bóng đá S giữa 2 đội A và B mà đội A quá mạnh, chắc chắn thắng đội B thì độ bất định của trận đấu S là bằng 0: không cần xem cũng biết kết quả. Ngược lại nếu 2 đội hoàn toàn cân tài cân sức, khả năng A thắng là 1/3, B thắng là 1/3, hai đội hoà là 1/3 thì khi đó tính bất định của trận đấu S mới là cao nhất!

2.2. Entropy và lượng thông tin – Đơn vị đo lượng thông tin

Trước khi tiến hành một phép thử S, ta chưa có "hiểu biết" gì về kết cục của quan sát đó, vì vậy độ bất định là H(S). Nhưng sau khi tiến hành xong quan sát S, kết cục của quan sát đã xảy ra: không còn tính bất định nữa, rõ ràng quan sát S đã mang lại một "cái gì đó" có tác dụng "khử", xoá bỏ độ bất định H)S). Ta nói rằng quan sát S đã mang lại một *lượng thông tin* có tác dụng xoá bỏ độ bất định H(S). Như vậy có thể hiểu là *thông tin* do một quan sát S mang lại đã "*khử*" độ bất định chứa trong quan sát đó. Độ bất định H(S) càng lớn thì khi khử được nó ta thu được một lượng thông tin càng lớn. *Thông tin không phải là độ bất định – nó trái ngược với độ bất định và dùng để "khử" độ bất định, nhưng về "độ lớn" thì lượng thông tin tỷ lệ thuận với độ bất định.* Nếu quan sát S chứa một độ bất định H(S) càng lớn thì khi xoá bỏ được H(S), ta đã thu được một lượng thông tin tương ứng càng lớn. Quan hệ giữa độ bất định và lương thông tin có thể so sánh tương tự như quan hệ giữa *trọng lực* và *lực* trong Vật lý học. Trong Vật lý học, hai khái niệm trọng lực và lực hoàn toàn khác nhau, trọng lực của một vật được xem là sức hút của tâm trái đất lên vật đó, có tác dụng làm cho vật ấy rơi tự do thẳng đứng với gia tốc g. Muốn "khử" tác động của trọng lực, ta phải dùng một lực có cùng độ lớn nhưng tác động ngược chiều, vì vậy trọng lực và lực có thể được "đo" bằng cùng một loại đơn vị. Cũng như vậy, độ bất định - entropy - trong một quan sát S và lượng thông tin do quan sát S mang lại là hoàn toàn khác nhau, đối lập nhau, lượng thông tin dùng để khử entropy, nhưng "độ lớn" của chúng về giá trị tuyệt đối là bằng nhau, vì vậy ta có thể dùng cùng một loại đơn vị để "đo lường" cả entropy và lượng thông tin.

Ta chọn làm đơn vị đo *Entropy* và lượng thông tin là Entropy của một quan sát nhị phân S có 2 kết cụ*c đồng khả năng* A_1, A_2 mà $p_1 = P(A_1) = p_2 = P(A_1) = ½$, khi đó:

$$H(S) = ½ \log_a ½ + 1/2\log_a ½ = \log_a 2 = 1$$
(*nếu dùng cơ số a =2*)

Đơn vị đó được gọi là đơn vị nhị phâ*n – binary unit* – viết tắt là *bit.*

Vì đơn vị bit khá bé nên trong thực tế ta thường dùng các bội số của bit như là:

1 Byte	= 103 bits
1 KiloByte	= 103 Bytes

1 MegaByte = 103 KiloBytes

1 GigaByte = 103 MegaBytes

1 TeraByte = 103 GigaBytes

2.3. Entropy có diều kiện và Thông tin tương hỗ.

Xét 2 phép thử: A có các kết cục có thể là A_i, i = {1, 2, …,m}với xác suất xuất hiện tương ứng là $p_i = P(A=A_i)$ và B có các kết cục Bj, j = {1,2,…n}với XS xuất hiện là $q_j = P(B=B_j)$. Phép thử đồng thời AB có các kết cục có thể là $(A=A_i, B=B_j)$ với XS xuất hiện tương ứng là $r_{ij} = P(A=A_i, B=B_j)$. nếu các phép thử A và B là độc lập thì: $r_{ij} = p_i.q_j$ còn trong trường hợp A và B có liên quan thì theo công thức xác suất có điều kiện: $r_{ij} = P(A_i, B_j) = P(A_i) P(B_j/A_i) = P(Bj).P(A_i/B_j)$.

Entropy của A, B và AB là: $H(A) = -\Sigma_i p_i \log p_i$; $H(B) = - \Sigma_j q_j \log q_j$ và $H(AB) = -\Sigma_i \Sigma_j r_{ij} \log r_{ij}$.

Trong trường hợp A và B độc lập, nghĩa là khi biết các kết cục của A hoặc B không ảnh hưởng gì đến nhau, theo công thức nhân XS độc lập ta có: H(AB) = H(A) * H(B).

Nhưng khi A và B không độc lập thì theo công thức XS có điều kiện:

H(AB) = H(A) * H(B/A) = H(B) * H(A/B).

Chẳng hạn khi gieo 1 con xúc xắc đối xứng, nếu xét A là quan sát: ra mặt mấy điểm thì H(A)=6, và B là quan sát: ra mặt chẵn hay lẻ: thì H(B) = 2. Nhưng nếu đã biết kết cục B là: ra mặt lẻ rồi thì khi đó H(A/điều kiện B lẻ) chỉ còn là 3. Nếu chưa có quan sát B thì entropy của A là H(A), nhưng *với điều kiện đã biết các kết cục của B* thì entropy về A chỉ còn lại là: $H(A/B) = - \Sigma_j q_j \Sigma_i PA_i/B_j) \leq H(A)$. Như vậy: *Tác dụng của kết cục của quan sát B đã làm giảm độ bất định về A một đại lượng là H(A) – H(A/B).*

Ta có định nghĩa: *Lượng thông tin về A do B mang lại là hiệu số giữa độ bất định toàn phần về A với độ bất định có điều kiện của A khi đã biết kết quả về B.*

$$I(A,B) = H(A) - H(A/B) \qquad (4)$$

Tương tự như vậy ta cũng có thể định nghĩa *lượng thông tin về B do A mang lại là hiệu số giữa độ bất định toàn phần về B với độ bất định có điều kiện của B khi đã biết kết quả về A:*

$$I(B,A) = H(B) - H(B/A) \qquad (5)$$

Dễ dàng chứng minh được là I(A,B) = I(B,A) cho nên ta gọi:

Lượng thông tin tương hỗ giữa 2 quan sát A và B là:

$$I(A,B) = I(B,A) = H(A) - HA/B) = H(B) - H(B/A) \qquad (5)$$

Rõ ràng:

$$0 \leq I(A,B) = I(B,A) \leq H(A) \text{ hoặc } H(B) \qquad (6)$$

Thí dụ 1: Gieo 1 con xúc xắc. Gọi A là quan sát số điểm của con xúc xắc, rõ ràng A có 6 kết cục và $H(A) = \log_a 6$; gọi B là quan sát: con xúc xắc có số điểm chẵn B_1:(2,4,6) hay số điểm lẻ B_2: (1,3,5) vậy B có 2 kết cục và $H(B) = \log_a 2$. Nhưng nếu đã có kết quả của quan sát B rồi – nghĩa là đã biết được con xúc xắc có điểm chẵn hay lẻ rồi – thì chỉ còn có thể có 3 kết cục về A, vì vậy: $H(A/B) = \log_a 3$

Như vậy: Sau khi đã có quan sát B thì độ bất định về quan sát A đã giảm đi một lượng là:

$$I(A,B) = H(A) - H(A/B) = \log_a 6 - \log_a 3 = \log_a(6/3) = \log_a 2$$

Thí dụ 2: Gọi A là quan sát số điểm khi gieo một con xúc xắc: $H(A) = \log_a 6$. C là quan sát kết quả khi rút 1 trong 52 quân bài tu lơ khơ: $H(C) = \log_a 52$. Nếu đã biết trước kết quả trong quan sát C là rút được quân bài nào trong 52 quân thì khi tiến hành gieo con xúc xắc ta vẫn có 6 kết cục đồng khả năng là {1,2,3,4,5,6} không có gì thay đổi.

Dễ dàng tính được: H(A/C) = H(A) và H(C/A) = H(C) cho nên: I(A,C) = I(C,A) = 0: Hai quan sát A và C hoàn toàn không có liên quan về mặt thông tin, nói khác đi: Khi biết kết quả của một quan sát này thì không ảnh hưởng gì đế độ bất định trong quan sát kia – điều này hoàn toàn phù hợp với ý nghĩa trực quan của 2 quan sát!

2.4. Khái niệm tỷ số thông tin

Như đã định nghĩa ở mục 3: Lượng thông tin tương hỗ I(A,B) được xem là: Thông tin về A do B mang lại, hoặc thông tin về B do A mang lại và:

$$0 \leq I(A,B) \leq H(A) \text{ hoặc } H(B)$$

• *Trường hợp giới hạn 1:*

$$I(A,B) = 0$$

Lượng thông tin thu được từ quan sát B không làm giảm độ bất định trong quan sát A tức là hai quan sát A, B độc lập về mặt thông tin, *quan sát B không mang lại thông tin nào về quan sát A.*

• *Trường hợp giới hạn 2:*

$$I(A,B) = H(A)$$

Lượng thông tin thu được từ quan sát B khử hết độ bất định trong quan sát A tức là *quan sát B mang lại thông tin toàn phần về quan sát A.*

Trong những trường hợp trung gian: I(A,B) càng lớn, càng gần bằng H(A) thì vai trò của quan sát B đối với quan sát A càng quan trọng.

Trong [8] I. Ibrahimovitch đã đưa ra một khái niệm mới: khái niệm về *tỷ số thông tin.*

Định nghĩa: Ta gọi tỷ số thông tin (*Information Ratio)* giữa quan sát B đối với quan sát A là

$$R(A/B) = \frac{I(A,B)}{H(A)} \qquad (7)$$

Tỷ số đó biểu thị mối quan hệ về mặt thông tin giữa một quan sát ngẫu nhiên A – thường gọi là *quan sát cốt yếu* – đối với một quan sát B – thường gọi là *quan sát sơ bộ*.

Rõ ràng là:

$$0 \leq R(A/B) \leq 1 \qquad (8)$$

• *Trường hợp giới hạn 1:*

R(A/B) = 0 kết quả của quan sát B không mang lại thông tin gì về quan sát A, quan sát B không có tác dụng gì đối với quan sát A.

• *Trường hợp giới hạn 2:*

R(A/B) = 1 kết quả của quan sát B mang lại đầy đủ thông tin về về quan sát A, quan sát B có tác dụng quyết định đối với quan sát A.

Trong những trường hợp trung giam, tỷ số R(A/B) càng lớn, càng gần giá trị 1 thì sự hiểu biết về A do B mang lại càng lớn, kết cục quan sát B mang lại càng nhiều thông tin cho kết cục của A.

III. Tỷ số thông tin và nghiên cứu mối liên quan giữa các yếu tố ngẫu nhiên phi định lượng

Trong toán học nói riêng và trong mọi lĩnh vực nghiên cứu khoa học tự nhiên, xã hội và con người nói chung ta thường phải xem xét mối liên quan giữa hai hoặc nhiều yếu tố ngẫu nhiên đồng thời xảy ra trên một tập hợp đối tượng nào đó. Trước hết ta xét trường hợp các yếu tố ngẫu nhiên cần khảo sát là những *yếu tố định lượng* được *(Quantitative Random Factor)* nghĩa là mỗi trạng thái, kết cục quan sát, có thể biểu thị bằng giá trị, bằng số của một đại lượng hay một biến số ngẫu nhiên. Chẳng hạn trong kinh tế học người ta cần nghiên cứu các mối liên quan giữa giá trị tổng thu nhập quốc gia của quốc gia đó Y với nhu cầu sử dụng điện năng X_1, nhu cầu lương thực thực phẩm X_2, nhu cầu đi lại trong và ngoài nước X_3, nhu cầu dịch vụ giải trí X_4 v..v.., Y và X_1, X_2, X_3, X_4… đều là những đại lượng, có thể có thể biểu thị bằng những con số kèm theo đơn vị đo, chẳng hạn Y *USD*, X_1 *MegaW*, X_2 *tấn*, X_3 *người.km* hay X_4 *USD*…Hay trong nghiên cứu y sinh học người ta cần nghiên cứu mối liên quan giữa các chỉ số chiều cao X (*mét*), cân nặng Y (*kg*), vòng ngực Z (*cm*), lượng hồng cầu trong máu T (đơn vị)… của một con người.

Để nghiên cứu mức độ chặt chẽ của các quan hệ giữa các yếu tố định lượng đã có một bộ công cụ toán học rất hoàn chỉnh trong Lý thuyết Xác suất và thống kê toán học: đó là *Lý thuyết tương quan và hồi qui - Correlation and regression theory.* [6]

Trong rất nhiều trường hợp khác, các yếu tố ngẫu nhiên cần quan sát không phải là những yếu tố định lượng mà chỉ là những yếu tố sắp thứ tự được. Chẳng hạn ta cần xét các năng khiếu của học sinh về âm nhạc A, hội hoạ B, văn học C, toán học D, ngoại ngữ E, thủ công F… có liên quan gì với nhau hay không? Phải chăng học sinh giỏi toán thì thường dốt văn mà giỏi thủ công thì thường hội hoạ cũng giỏi? Để phân loại trình độ về âm nhạc, văn học, … cho một tập hợp học sinh thường người ta khảo sát qua các cách thi cử và cho điểm. Tuy nhiên cần thấy rằng: Điểm số của một học sinh là một cách xếp thứ tự chứ hoàn toàn không phải là một đại lượng! Vì rằng: Một đại lượng thì phải có tính cộng được – *addability* - và có thể chọn ra một đơn vị đo (*unit)* cho đại lượng đó. Chẳng hạn sản lượng lúa của một cánh đồng là đại lượng có đơn vị là Tấn (hoặc kg …), tổng sản lượng của cánh đồng có 3 thửa ruộng có sản lượng lần lượt là 3T, 2T và 10T cũng bàng đúng bằng tổng sản lượng của cánh đồng có 3 thửa mà mỗi thửa đều có sản lượng 5T. Trong khi đó nếu xét trình độ môn Toán của một 3 học sinh qua 3 bài thi có điểm thi là 3, 2, 10 điểm và một học sinh cả 3 kỳ thi đều điểm 5 thì hoàn toàn khác nhau! Mặt khác: Không thể nào chọn ra được một đơn vị đo năng khiếu cho học sinh đối với bất kỳ môn học nào cả!

Để nghiên cứu mối liên quan giữa các yếu tố ngẫu nhiên không định lượng mà là yếu tố ngẫu nhiên sắp thứ tự được, nhiều nhà toán học như Spearman, Kendall cũng đã đề xuất một công cụ có hiệu lực đang được sử dụng rộng rãi là lý thuyết về tương quan hạng hay tương quan thứ tự. *Rank Corelation* - [7]. Tuy nhiên trong thực tế, ngoài các yếu tố định lượng được và yếu tố sắp thứ tự được còn có những yếu tố *hoàn toàn không định lượng được* mà cũng *không sắp thứ tự được,* ví dụ như có thể đặt vấn đề: Năng khiếu con người có phụ thuộc địa phương gốc gác không? Phải chăng người miền này thì giỏi thơ ca, người miền nọ lại xuất hiện nhiều tài năng toán học? Hay là trong vấn đề nghiên cứu thị hiếu thị trường, người ta thường cảm thấy chắc chắn là có một mối liên quan nào đó giữa các loại hàng hoá đối với dân cư các vùng miền, các ngành nghề trong xã hội. Chẳng hạn, dân cư các vùng miền khác nhau trong một nước rõ ràng là có thị hiếu về hàng tiêu dùng ăn, mặc v..v..

khác nhau, nhưng mức độ khác nhau là thế nào? Nên phân loại ra những vùng miền như thế nào đối với mỗi loại hàng hoá? Nghiên cứu được qui luật liên quan chính xác giữa các yếu tố đó góp phần rất quan trọng trong hiệu quả giáo dục đào tạo, kinh doanh v..v..

Như đã nói trên các yếu tố như vùng miền, nghề nghiệp, năng khiếu, thị hiếu v..v.. Đều không định lượng được mà cũng không sắp thứ tự được: Vì vậy các công cụ toán học về tương quan, hồi qui và tương quan hạng v..v.. đều không thể áp dụng được. Khái niệm tỷ số thông tin mở ra một hướng nghiên cứu chính xác và đầy triển vọng cho việc nghiên cứu các yếu tố ngẫu nhiên dạng không thể lượng hoá và không thể sắp thứ tự này trong thiên nhiên, xã hôi và con người. Một hiện tượng X trong tự nhiên, xã hội hay con người luôn chứa một độ bất định được "đo" bằng *entropy H(X)*. Muốn tìm hiểu về *quan sát cốt yếu X* ta có thể tiến hành những *quan sát* sơ bộ khác chẳng hạn như S để thu được một lượng thông tin về X: I(S, X) = H(X) – H(X/S). Nếu lượng thông tin thu được I(S, X) càng gần với H(X) thì độ bất định về X càng giảm sau khi làm quan sát S, nghĩa là *quan sát sơ bộ S càng có ý nghĩa khi nghiên cứu quan sát cốt yếu X*.

Tỷ số thông tin của quan sát S đối với quan sát X là:

$$R(S, X) = \frac{I(S,X)}{H(X)} = \frac{H(X)-H(X/S)}{H(X)}$$

Và: $0 \le R(S,X) \le 1$

Tỷ số thông tin R(S,X) có thể dùng làm một độ đo mức độ liên quan về mặt thông tin giữa quan sát S đối với quan sát X *trong trường hợp trong các quan sát X và S là có thể có những yếu tố ngẫu nhiên không lượng hoá mà cũng không sắp thứ tự được*.

IV. Kết luận

Lý thuyết thông tin là những nghiên cứu toán học về định lượng, lưu trữ và truyền tải thông thông tin. C.E. Shannon cung cấp một định nghĩa toán học về thông tin và mô tả chính xác lượng thông tin có thể được truyền đạt giữa các thành tố khác nhau của bất kỳ hệ thống nào. Thông tin được coi là một tập hợp các thông điệp, được thu thập, lữu trữ dưới các dạng thức vật lý khác nhau, mục tiêu là để nhận định, đánh giá độ bất định của thông tin, đặc biệt những thông tin không định lượng. Khái niệm tỷ số thông tin có khả năng phát huy tác dụng rất to lớn trong vấn đề nghiên cứu mối quan hệ giữa các yếu tố không định lượng của các đối tượng trong thế giới tự nhiên, xã hội và con người, đặc biệt là trong các vấn đề năng khiếu trong giáo dục cũng như vấn đề thị hiếu trong nghiên cứu thị trường mà chúng ta sẽ có dịp đề cập đến trong những đề tài nghiên cứu tiếp sau.

Tài liệu tham khảo:

[1]. Thái Thanh Sơn. Giáo trình Lý thuyết Thông tin và Ứng dụng, ĐHBK, 2000.

[2]. Thái Thanh Sơn và cộng sự. Tin học cơ bản và ứng dụng trong đời sống, NXB Thông tin và Truyền thông, 2014.

[3]. Claude E. Shannon. A mathemathical model of Information. MIT, 1948.

[4]. A. Feinstein. Information theory and its applications. MIT, 1954.

[5]. George N. Saridis. Entropy in Control Engineering, World Scientifics, 2001

[6]. A.Yaglom. Correlation Theory of Stationary and Related Random functions, Springer, 1987

[7]. Kendal M.G. Rank Corelation Method - London: Griffin, 1970. ISBN 0-85264-199-0

[8]. И. Ибрагимович. Теория информации и ее приложения, Москва, 1986

THE SHANNON MATHEMATICAL MODEL OF INFORMATION AND THE INFORMATION RATIO

***Thai Thanh Son*†, *Thai Thanh Tung*†**

Abstract: *In the last half-century, the term Information - has been used in a widespread way all the time, all over the world. From technical scientists to ordinary people, they talk about information daily, care about it, and always use it in all areas of their lives. However, it is not easy to give an exact "definition" for that concept. In this paper, we briefly and fully present Shannon's mathematical model and provide specific illustrations to assist those initially learning about Information Technology in accessing the most important concept in the field. In particular, we then reviewed the concept of the information ratio between two observations, which plays a very important role in investigating the relationship between non-quantitative random events. This research field has yet to be fully exploited.*

Keywords: *information, information uncertainty, communication information theory, information ratio.*

† Hanoi Open University