Entropia de Shannon

Autor:
Data de publicació: 16-07-2018

L'entropia de Shannon , a causa de Claude Shannon , és una funció matemàtica que, intuïtivament, correspon a la quantitat d' informació continguda o lliurada per una font d'informació. Aquesta font pot ser un text escrit en un idioma determinat, un senyal elèctric o qualsevol fitxer informàtic (col · lecció de bytes). Des del punt de vista d'un receptor, com més emeti informació diferent, més entropia (o incertesa sobre el que emet la font) és gran, i viceversa. Com més receptor rep informació sobre el missatge transmès, més entropia (incertesa) respecte a aquest missatge disminueix, a la llum d'aquest guany d'informació. La definició d'entropia de Shannon és tal que com més redundant sigui la font, menys informació conté. En absència de restriccions particulars, l'entropia és màxima per a una font de la qual tots els símbols són equiprobables.

En el cas particular d'un sistema de telecomunicacions , l'entropia de la font d'informació (el transmissor) indica la incertesa del receptor respecte del que la font transmetrà. Per exemple, una font que es considera que sempre envia el mateix símbol, per exemple, la lletra 'a', té zero entropia, és a dir, mínima. De fet, un receptor que coneix només les estadístiques de transmissió de la font assegura que el símbol següent serà un "a", sense que ningú no s'equivocarà. El receptor no necessita rebre un senyal per eliminar la incertesa sobre el que ha transmès la font perquè no genera cap perill. D'altra banda, si es considera que la font envia un 'a' meitat de temps i a " " l'altra meitat, el receptor no està segur de la següent carta a rebre. L'entropia de la font en aquest cas, per tant, no és zero (positiu) i representa quantitativament la incertesa que reina sobre la informació que emana de la font. Des del punt de vista del receptor, l'entropia indica la quantitat d'informació que necessita obtenir per eliminar completament la incertesa (o dubte) sobre el que ha transmès la font.

resum
[masquer]

1Història
2Preàmbul
3Entropia d'un text comú
4Definició formal

4.1Justificació del logaritme
4.2Maximització de l'entropia

4.2.1Desigualtat de Gibbs
4.2.2Demostracions

4.2.2.1Evidència de la desigualtat de Jensen
4.2.2.2Prova per un terminal lineal en el logaritme

5Propietats
6Útil pràctica

7exemples simples

7.1Urns
7.2Text

8Vegeu també
9Bibliografia

història
El 1948, mentre treballava a Bell Laboratories , l'enginyer elèctric Claude Shannon va formalitzar matemàticament la naturalesa estadística de la "informació perduda" en senyals de línia telefònica. Per això, va desenvolupar el concepte general d'entropia d'informació, fonamental en la teoria de la informació 1 .

Inicialment, no sembla que Shannon tingués especial coneixement de l'estreta relació entre la seva nova mesura i el treball anterior en termodinàmica . El terme entropia va ser suggerit pel matemàtic John von Neumann per la raó que aquesta noció s'assemblava a la que ja es coneixia com entropia en física estadística, i hauria afegit que aquest terme es va entendre malament per triomfar qualsevol debat 2 .

El 1957 , Edwin Thompson Jaynes demostrarà el vincle formal entre l' entropia macroscòpica introduïda per Clausius el 1847, el microscòpic introduït per Gibbs i l'entropia matemàtica de Shannon. Aquest descobriment va ser descrit per Myron Tribus com una "revolució passada desapercebuda" 3 .

preàmbul
A principis de la dècada de 1940, les telecomunicacions estaven dominades per analògics . La transmissió de ràdio i televisió es van basar en modulacions contínues com lamodulació d'amplitud (AM) i la modulació de freqüència (FM). Els sons i les imatges es transformen en senyals elèctrics l'amplitud i / o freqüència dels quals són funcions contínues, de vegades proporcionals, al senyal d'entrada. En el cas del so, es mesura amb un micròfon el fenomen de la pressió i la depressió que viatgen a l'aire. En el cas de la televisió, la blancor de la imatge (la seva brillantor) és el principal senyal d'interès. Aquest procediment implica que el soroll afegit durant la transmissió produeix una degradació de la senyal rebuda. L'arquetip d'aquest tipus de soroll pren la forma de xerraire de ràdio i neu per a la televisió. La modulació analògica implica l'ús de nombres reals la dilatació decimal és infinita per representar informació (pressió sonora, intensitat de la llum, etc.). Un soroll, per petit que sigui, tingui una conseqüència directa en el senyal.

Els investigadors han admès així que una manera eficaç de protegir el soroll seria transformar el so i la imatge en nombres discrets, en lloc d'utilitzar nombres reals la precisió dels quals requereix un nombre infinit de dígits. Per exemple, es podria utilitzar el número 0 per representar la negrit total i el número 10 per a un blanc perfecte, amb tots els enters entre els dos que representen nivells successius de gris. Si 11 nivells no semblen suficients, podem utilitzar el mateix mètode per a una divisió d'intensitats tan grans com sigui necessari per satisfer l'ull. Es pot fer un raonament similar per al so i arribem a un punt on és possible representar una pel·lícula i la seva banda sonora amb una quantitat limitada d'enters.

La transmissió d'aquests enters provoca el que anomenem comunicació digital. Si el soroll que parlem en el cas analògic es considera en una transmissió digital, es produiran errors quan aquest soroll és prou fort com per convertir un número en un altre. En el cas analògic, fins i tot un petit soroll es converteix en errors perceptibles. En digital, és poc probable que es produeixi un petit soroll per generar un error, però el soroll pot, però, fer-ho. Els investigadors han pensat que un ha d'acceptar acceptar que la comunicació perfecta era impossible. És aquesta conjectura que Shannon va ser refutar per la seva teoria de la informació. Va haver de demostrar que era possible transmetre informació sense errors utilitzant una estratègia de codificació digital mentre estiguéssim contents amb una determinada velocitat de transmissió. Per error aquí, significa la capacitat del receptor per restaurar el missatge original fins i tot si el missatge rebut és modificat pel soroll.

L'entropia de Shannon, mesura del contingut informatiu d'un missatge, s'uneix als teoremes de Shannon per decidir el ràpid que volem si volem tenir l'esperança de transmetre les dades d'aquest missatge sense cap error. No cal dir que un soroll més potent distorsiona encara més un missatge transmès i Shannon prediu que, en presència d'un soroll més gran, cal reduir la velocitat de transmissió per arribar al mateix resultat sense cap error. Una estratègia de codificació elemental i històricament utilitzada en telegrafia és la col·lació o la transmissió múltiple (generalment doble) de la mateixa informació. De fet, la probabilitat d'obtenir errors en la majoria d'aquesta informació és menor que la probabilitat

d'obtenir un error d'una sola transmissió. Una transmissió per triplicat permetria a un sistema de votació veure on es troba l'anomalia, incloent-hi la falta deredundància del codi (per exemple, la transmissió de números de part a l'ordre en una nomenclatura). No obstant això, aquesta és una codificació ingènua i no permet assolir els límits que planteja Shannon.

El càlcul de l'entropia d'una font de missatge proporciona una mesura de la informació mínima que hem de mantenir per representar aquestes dades sense pèrdua. En termes comuns, això significa per al cas particular de compressió d'arxius d'ordinador que l'entropia indica el nombre mínim de bits que un arxiu comprimit pot arribar. S'ha d'entendre que si estem disposats a perdre dades, com quan es comprimeixen els sons per format MP3 o quan es comprimeixen imatges per vídeos JPEG o MPEG, podem creuar aquest límit inferior imposat pel entropia de la imatge original. En realitat, primer vam baixar l'entropia de la imatge o el so eliminant detalls imperceptibles per als humans. La nova entropia reduïda és llavors el nou límit inferior per a la posterior compressió sense pèrdua.

Entropia d'un text comú
Shannon ofereix una manera senzilla de determinar l'entropia d'un text donat per a un receptor determinat 4 : A té el text i demana a B que endevinem lletra per lletra (espais inclosos). si B endevina correctament la lletra, es compta 1 (perquè A, mentre respon "sí" a ell, li va transmetre 1 bit d'informació). Si B s'equivoca, se li dóna la lletra correcta i compta 4.75 (perquè un caràcter de 26 (és a dir, 27 - 1) representa 4.75 bits d'informació).

El mètode aplicat als textos dels diaris i als lectors actuals mostra que es pot endevinar una lletra de dos, la redundància del llenguatge actual era, per tant, un factor de 2, i el contingut informatiu d'una lletra en aquest context Només 2,35 bits.

Aquesta senzilla mesura està ocupada per Léon Brillouin en el seu llibre Science and Theory of Information .

Definició formal
Per una font, que és una variable aleatòria discreta, { displaystyle X} que comprèn { displaystyle n} símbols, símbol { displaystyle i} tenint una probabilitat { displaystyle P_ {i}} aparèixer, entropia { displaystyle H} des de la font { displaystyle X} es defineix com:

{ displaystyle H_ {b} (X) = - mathbf {E} [log_ {b} {P (X = x_ {i})} = sum _ {i = 1} ^ {n} P_ {i} log _ {b} left ({ frac {1} {P_ {i}}} right) = - sum _ {i = 1} ^ {n} P_ {i} log _ { b} P_ {i}. , !}
on { displaystyle mathbf {E}} denota l' expectativa matemàtica . Normalment s'utilitza un logaritme basat en 2 perquè l'entropia té les unitats bit / symbol. Els símbols representen les possibles realitzacions de la variable aleatòria X.

{ displaystyle H (X) = H_ {2} (X) = - sum _ {i = 1} ^ {n} P_ {i} log _ {2} P_ {i}. , !
L'entropia així definida verifica les següents propietats:

{ displaystyle H (X) geq 0} amb igualtat ssi { displaystyle exists i | P (X = x_ {i}) = 1}

és major o igual que zero, amb igualtat per a una distribució resumida en un punt, és a dir, zero en tots els aspectes { displaystyle i} excepte en un punt { displaystyle i *} per a això { displaystyle p (i *) = 1} ;

{ displaystyle H (X) leq log (n)}

és màxim per a una distribució uniforme, és a dir, quan tots els estats tenen la mateixa probabilitat;
Totes les coses són iguals, augmenta amb el nombre d'estats possibles (el que tradueix la intuïció que com més opcions hi ha, més gran serà la incertesa);

ella és contínua

Justificació del logaritme
L'ús del logaritme pot semblar arbitrari a primera vista. Cal recordar que l'efecte del logaritme és transformar la multiplicació en suma i la divisió en la resta. A més, donades dues variables aleatòries independents { displaystyle X} i { displaystyle Y} , la probabilitat del producte cartesià d'aquestes variables aleatòries ve donada per:

{ displaystyle P left (X, Y right) = P left (X right) P left (Y right)}

per tant:

{ displaystyle H (XY) = - sum_ {x in X} sum_ {y in Y} P (x, y) log P (x, y) = - sum _ {x in X} sum _ {y in Y} P (x) P (y) log P (x) P (y)} { displaystyle Rightarrow H (XY) = - sum _ {x in X} sum_ {y in Y} P (x) P (y) left [log P (x) + log P (i) right]} { displaystyle Rightarrow H (XY) = - sum _ {x in X} sum_ {y in Y} P (x) P (y) log P (x) - sum _ {y en Y} sum _ {x in X} P (x) P (y) log P (y)} { displaystyle Rightarrow H (XY) = - sum _ {x in X} P (x) log P (x) sum _ {y in Y} P (y) - sum _ {y en Y} P (y) log P (i) sum _ {x in X} P (x)} { displaystyle Rightarrow H (XY) = - sum_ {x in X} P (x) log P (x) - sum_ {y in Y} P (y) log P (y) } { displaystyle Rightarrow H (XY) = H (X) + H (Y)}
Que és intuïtivament satisfactori des de llavors { displaystyle X} i { displaystyle Y} són independents. Atès que l'entropia és una mesura d'informació mitjana continguda en una variable aleatòria, l'entropia de la combinació de dues variables no relacionades només s'afegeix. A continuació, veiem el treball realitzat pel logaritme fent el pont entre la multiplicació de probabilitats i l'addició d'entropies. En cas que hi hagi una certa dependència entre { displaystyle X} i { displaystyle Y} es verifica per proves similars que:

{ displaystyle H left (XY right) = H left (X right) + H left (Y | X right)}
Maximització de l'entropia
Desigualtat de Gibbs
La proposició anterior, segons la qual una distribució d'esdeveniments equiprobables maximitza l'entropia per a una determinada quantitat d'elements { displaystyle K} , es pot expressar de forma més general per la desigualtat de Gibbs:

{ displaystyle H (X) leq - sum _ {i} p_ {i} log q_ {i}} { displaystyle sum _ {i} p_ {i} log q_ {i} leq sum _ {i} p_ {i} log p_ {i}}
on { displaystyle q_ {i}} és qualsevol distribució de probabilitat en la variable X. Vegem llavors que la desigualtat precedent s'obté com un cas especial quan { displaystyle 1 / q_ {i} = K} , és a dir, per esdeveniments igualment probables:

{ displaystyle H (X) leq - sum _ {i} p_ {i} log q_ {i} = sum _ {i} p_ {i} log { frac {1} {q_ {i} }} = sum _ {i} p_ {i} log K = log K}
demostracions
Evidència de la desigualtat de Jensen
El logaritme és una funció còncava, és a dir, la seva segona derivada és menor o igual que zero per a qualsevol valor de { displaystyle x} en el seu domini. Per Jensen obtenim llavors:

{ displaystyle E [f (X)] leq f (E [X])}
A continuació, pregunta:

{ displaystyle x = { frac {q_ {i}} {p_ {i}}}} { displaystyle f left (x right) = log left (x right)}
Substitució a Jensen:

{ displaystyle E left [log { frac {q_ {i}} {p_ {i}}} right] leq log left (E left [{ frac {q_ {i}} {p_ {i}}} right] right)}
I desenvolupant expectatives matemàtiques:

{ displaystyle sum_ {i} p_ {i} log { frac {q_ {i}} {p_ {i}}} leq log sum _ {i} p_ {i} { frac {q_ {i}} {p_ {i}}} = log sum _ {i} q_ {i} = log 1 = 0} { displaystyle Rightarrow sum _ {i} p {i} log { frac {q_ {i}} {p_ {i}}} leq 0}
Per la propietat dels logaritmes:

{ displaystyle sum _ {i} p_ {i} log q_ {i} leq sum _ {i} p_ {i} log p_ {i} = - H (X)} { displaystyle Rightarrow H (X) leq - sum _ {i} p_ {i} log q_ {i}}
QED.

Prova d'un enllaç lineal sobre el logaritme
És fàcil verificar que la funció logarítmica està delimitada per qualsevol línia tangent a ella. La naturalesa còncava de la funció del logaritme li dóna aquesta propietat:

{ displaystyle log (z) leq z-1}
Així tenim:

{ displaystyle sum_ {i} p_ {i} log { frac {q_ {i}} {p_ {i}}} leq sum _ {i} p_ {i} left [{ frac { q_ {i}} {p_ {i}}} - 1 right] = sum _ {i} left [q_ {i} -p_ {i} right] = sum _ {i} q_ {i} - sum _ {i} p_ {i} = 1-1 = 0} { displaystyle Rightarrow sum _ {i} p {i} log { frac {q_ {i}} {p_ {i}}} leq 0}

El final de la prova és el mateix que abans, per la propietat dels logaritmes:

{ displaystyle sum _ {i} p_ {i} log q_ {i} leq sum _ {i} p_ {i} log p_ {i} = - H (X)} { displaystyle Rightarrow H (X) leq - sum _ {i} p_ {i} log q_ {i}}
QED.

propietats
Aquí teniu algunes propietats importants de l'entropia de Shannon:

{ displaystyle H (X) geq H (X | Y)}
{ displaystyle H (X | Y) geq H (X | YZ)}
{ displaystyle H (X_ {1} ldots X_ {n}) = H (X_ {1}) + H (X_ {2} | X_ {1}) + ldots + H (X_ {n} | X_ { 1} ldots X_ {n-1})}
{ displaystyle H (X_ {1} ldots X_ {n}) leq sum _ {i = 1} ^ {n} H (X_ {i})}

Utilitat pràctica
L'entropia de Shannon s'utilitza en l'electrònica digital per digitalitzar una font utilitzant el màxim nombre possible de bits sense perdre informació. També quantifica el nombre mínim de bits en què es pot codificar un fitxer, mesurant així els límits que els algorismes de compressió sense pèrdua poden esperar aconseguir. També s'utilitza en altres camps, com, per exemple, per seleccionar el millor punt de vista d'un objecte tridimensional 5 .

Exemples simples
urnes
Penseu en la possibilitat d'una urna que conté diverses boles de diferents colors, des d'on s'agafa una bola a l'atzar (amb descompte). Si totes les boles tenen colors diferents, la nostra incertesa sobre el resultat d'un sorteig és màxima. En particular, si haguéssim d'apostar pel resultat d'un empat, no podríem afavorir una altra opció. D'altra banda, si un color determinat és més representat que els altres (per exemple, si l'urna conté més boles vermelles), la nostra incertesa es redueix lleugerament: la bola triada és més probable que sigui vermella. Si haguéssim d'apostar pel resultat d'un empat, apostaríem per una bola vermella. D'aquesta manera, el fet de revelar el resultat d'un sorteig proporciona, en general, més informació en el primer cas que en el segon, perquè l'entropia del "senyal" (computable des de la distribució estadística) és més alta.

text
Agafem un altre exemple: considerar un text en francès codificat com una cadena de lletres, espais i puntuacions (el nostre senyal és, doncs, una cadena de caràcters). Com que la freqüència d'alguns caràcters no és molt important (per exemple, 'w'), mentre que d'altres són molt comuns (per exemple, 'e'), la cadena de caràcters no és tan aleatòria. D'altra banda, sempre que no puguem preveure quin és el següent personatge, en certa manera, aquesta cadena és aleatòria, que la noció d'entropia de Shannon busca quantificar.

Vegeu també

Complexitat de Kolmogorov
Entropia de Renyi
Entropia mètrica
Entropia diferencial
Informació mútua
Teoria de la informació
El teorema de codi font

bibliografia

Claude E. Shannon Una teoria matemàtica de la comunicació Bell System Technical Journal , vol. [archive] 27, pp. [archive] 379-423 i 623-656, juliol i octubre de 1948 [archive]
? Sr. Tribus, EC McIrvine, "Energia i informació", Scientific American, 224 (setembre 1971).
? La referència es troba en aquest document [archive] (PDF)
? El mesurament depèn de la "cultura" del receptor. Una frase com "obtenim la següent sèrie ràpidament convergent" proporcionarà una taxa d'èxit més gran per als matemàtics que per als no matemàtics. De la mateixa manera, explica Brillouin, si s'utilitzen altres vocabularis molt especialitzats com mèdics, financers, polítics, etc.

? (en) P.-P. Vázquez, M. Feixas, M. Sbert, W. Heidrich, Selecció de punts de vista mitjançant entropia de visió , Actes de la Conferència sobre modelització i visualització de visions, 273-280, 2001.