48

What is the computer science definition of entropy?

computer-science information-theory entropy data-compression

I've recently started a course on data compression at my university. However, I find the use of the term "entropy" as it applies to computer science rather ambiguous. As far as I can tell, it roughly translates to the "randomness" of a system or structure.

What is the proper definition of computer science "entropy"?

share improve this question

fluffels
2,249 • 5 • 26 • 45

asked
Feb 4 '09 at 7:36

Zach Scrivena
20.7k • 7 • 47 • 61

edited
Feb 7 '09 at 17:02

Entropy is opening your clothes dryer and not finding the clothes already folded for you inside – **Brad Thomas** Feb 5 '16 at 15:13

add a comment

14 Answers

order by votes

44

Entropy can mean different things:

Computing

> In computing, entropy is the randomness collected by an operating system or application for use in cryptography or other uses that require random data. This randomness is often collected from hardware sources, either pre-existing ones such as mouse movements or specially provided randomness generators.

Information theory

> In information theory, entropy is a measure of the uncertainty associated with a random variable. The term by itself in this context usually refers to the Shannon entropy, which quantifies, in the sense of an expected value, the information contained in a message, usually in units such as bits. Equivalently, the Shannon entropy is a measure of the average information content one is missing when one does not know the value of the random variable

Entropy in data compression

Entropy in data compression may denote the randomness of the data that you are inputing to the compression algorithm. The more the entropy, the lesser the compression ratio. That

means the more random the text is, the lesser you can compress it.

> Shannon's entropy represents an absolute limit on the best possible lossless compression of any communication: treating messages to be encoded as a sequence of independent and identically-distributed random variables, Shannon's source coding theorem shows that, in the limit, the average length of the shortest possible representation to encode the messages in a given alphabet is their entropy divided by the logarithm of the number of symbols in the target alphabet.

share improve this answer

Niyaz
23.7k • 46 • 122 • 166

answered
Feb 4 '09 at 7:41

5 Actually, those are three statements of the same thing. – **Charlie Martin** Feb 4 '09 at 7:43

1 Yes, and that thing is called entropy, which is why it's ambiguous. – **fluffels** Feb 4 '09 at 7:49

1 Also, if those blocks are quoted, you should probably reference them. – **fluffels** Feb 4 '09 at 7:51

1 -1 for not providing reference. – **krosenvold** Feb 4 '09 at 7:54

1 Reference is here : Compression, and Information Content – **Isalamon** Mar 15 '13 at 21:03

show **1** more comment

11

My favorite definition, with a more practical focus, is found in Chapter 1 of the excellent book The Pragmatic Programmer: From Journeyman to Master by Andrew Hunt and David Thomas:

> **Software Entropy**
>
> While software development is immune from almost all physical laws, entropy hits us hard. Entropy is a term from physics that refers to the amount of "disorder" in a system. Unfortunately, the laws of thermodynamics guarantee that the entropy in the universe tends toward a maximum. When disorder increases in software, programmers call it "software rot."
>
> There are many factors that can contribute to software rot. The most important one seems to be the psychology, or culture, at work on a project. Even if you are a team of one, your project's psychology can be a very delicate thing. Despite the best laid plans and the best people, a project can still experience ruin and decay during its lifetime. Yet there are other projects that, despite enormous difficulties and constant setbacks, successfully fight nature's tendency toward disorder and manage to come out pretty well.
>
> ...
>
> ...
>
> **A broken window.**
>
> One broken window, left unrepaired for any substantial length of time, instills in the

> inhabitants of the building a sense of abandonment—a sense that the powers that be don't care about the building. So another window gets broken. People start littering. Graffiti appears. Serious structural damage begins. In a relatively short space of time, the building becomes damaged beyond the owner's desire to fix it, and the sense of abandonment becomes reality.
>
> The "Broken Window Theory" has inspired police departments in New York and other major cities to crack down on the small stuff in order to keep out the big stuff. It works: keeping on top of broken windows, graffiti, and other small infractions has reduced the serious crime level.
>
> Tip 4
>
> Don't Live with Broken Windows
>
> Don't leave "broken windows" (bad designs, wrong decisions, or poor code) unrepaired. Fix each one as soon as it is discovered. If there is insufficient time to fix it properly, then board it up. Perhaps you can comment out the offending code, or display a "Not Implemented" message, or substitute dummy data instead. Take some action to prevent further damage and to show that you're on top of the situation.

Text taken from: http://pragprog.com/the-pragmatic-programmer/extracts/software-entropy

share improve this answer

Ash
41k • 25 • 125 • 159

answered
Feb 4 '09 at 7:55

nathanjosiah
2,661 • 2 • 25 • 36

edited
Mar 17 '12 at 8:20

I knew I'd heard that somewhere.. – **Fortyrunner** Feb 4 '09 at 8:10

Yes, the broken window analogy is very good. – **Ash** Feb 4 '09 at 8:14

1 I'm pretty well certain that is only vaguely related to the question asked, though. Code entropy is only very slightly more rigorous than using the word 'entropy' as a metaphor. – **Charlie Martin** Feb 4 '09 at 15:34

@Charlie, Disagree, it is absolutely related to the question. "I find the use of the term "entropy" as it applies to computer science rather ambiguous". In CS, there are specialist definitions of entropy as well as a more general definition this answer provides. Hence fluffels question/confusion. – **Ash** Feb 5 '09 at 0:17

Starting to read, I didn't know I'll end up upvoting. This is very true especially with with projects that get new contributors. Usually inexperienced contributor follows previous design. So bad design is repeated. – **akostadinov** Jun 22 '15 at 14:01

add a comment

8

I always encountered entropy in the sense of Shannon Entropy.

From http://en.wikipedia.org/wiki/Information_entropy:

In information theory, entropy is a measure of the uncertainty associated with a random variable. The term by itself in this context usually refers to the Shannon entropy, which quantifies, in the sense of an expected value, the information contained in a message, usually in units such as bits. Equivalently, the Shannon entropy is a measure of the average

information content one is missing when one does not know the value of the random variable.

share improve this answer

Adrian Grigore
20.4k • 26 • 111 • 192

answered
Feb 4 '09 at 7:41

8

alt text

from University of Mexico

> The information theoretic notion of Entropy is a generalization of the physical notion. There are many ways to describe Entropy. It is a measure of the randomness of a random variable. It is also a measure of the amount of information a random variable or stochastic process contains. It is also a lower bound on the amount a message can be compressed. And finally it is the average number of yes/no questions that need to be asked about an random entity to determine its value.

Equation for Entropy in a sample application for probability calculation:

> it is the sum over all values of a rv of the probability of that value times the log of that prob(i.e. p(x)logp(x)). This equation can be derived from first principles of the properties of information.

share improve this answer

Ric Tokyo
5,776 • 2 • 21 • 44

answered
Feb 4 '09 at 7:42

ShreevatsaR
26.7k • 12 • 75 • 104

edited
Feb 4 '09 at 13:14

You should probably indicate that your second paragraph is a quote. – **fluffels** Feb 4 '09 at 7:52

Nit picking. In the final quote, shouldn't it say "times minus the log of that prob (base 2) (i.e. -p(x)log(p(x)))" In other words, information of each value, averaged over the values. – **Mike Dunlavey** Apr 30 '09 at 14:37

add a comment

3

In terms of compression and information theory, the entropy of a source is the average amount of information (in bits) that symbols from the source can convey. Informally speaking, the more unlikely a symbol is, the more surprise its appearance brings.

If your source has two symbols, say `A` and `B`, and they are equally likely, then each symbol conveys the same amount of information (one bit). A source with four equally likely symbols conveys two bits per symbol.

For a more interesting example, if your source has three symbols, `A`, `B`, and `C`, where the first two are twice as likely as the third, then the third is more surprising but is also less likely. There's a net entropy of 1.52 for this source, as calculated below.

You calculate entropy as the "average surprise", where the "surprise" for each symbol is its probability times the negative binary log of the probability:

```
                          binary
symbol  weight  probability   log    surprise
  A        2        0.4      -1.32    0.53
  B        2        0.4      -1.32    0.53
  C        1        0.2      -2.32    0.46
total      5        1.0                1.52
```

The negative of the binary log is used (of course) because logs of values between 0 and 1 (exclusive) are negative.

share improve this answer

joel.neely
22.8k • 8 • 43 • 60

answered
Feb 7 '09 at 17:42

2

http://en.wikipedia.org/wiki/Entropy_(information_theory)

share improve this answer

Ray Tayek
5,831 • 2 • 20 • 42

answered
Feb 4 '09 at 7:40

casperOne
59.5k • 10 • 134 • 205

edited
Feb 4 '09 at 8:01

2

Super SIMPLE definition

The word entropy can be defined in one sentence:

"The amount of information needed to describe a system."

Imagine for an example the expansion of the universe: From the beginning, all matter was collected in a small point before the big bang, so we could have described the system with "all matter is within one point." While today significantly more information is required to describe the system (the Universe, that is), one would need to describe all planetary positions, their movement, what's on them etc.. In terms of information theory, the definition

also works: E.g: The more letters you add to a password (the system), the more information is needed to describe the password. Then you can measure it in different units, eg bits or characters, like "hello" = 5 characters entropy = 40 bits of entropy (if charsize is 8 bits).

From this also comes that the more information you have the more ways you can arrange that information in. If you have 40 bits there are 2^40 different ways they can be arranged. If we are talking passwords here then the more possible arrangements of the information (bits) the longer it is going to take cracking (with brute force or dictionary attacks).

share improve this answer

Axel K
128 • 3 • 9

answered
May 19 '14 at 20:07

edited
May 19 '14 at 20:13

2

Here is a great alternate explanation for *entropy* in information theory.

> Entropy is a measure of **uncertainty involved in making a prediction**.

We can also describe entropy as how *surprised* we would be if we get an outcome after we made our initial prediction.

Let say we have a bent coin that gives us a head 99% of the time and a tail 1% of the time. Since there is only a one percent chance of getting a tail, we would be very surprised if we actually get a tail. On the other hand, it won't be too surprising if we got a head as we already have a 99 percent chance of getting a head.

lets assume that we have a function called `Surprise(x)` that would give us the amount of surprise for each outcome; then we can average the amount of surprise on a probability distribution. This average amount of surprise could also be used as a measure for how uncertain we are. This uncertainty is called ***entropy***.

Reference: this post.

share improve this answer

Greenish
1,072 • 1 • 14 • 24

answered
Jan 17 '16 at 18:07

0

Entropy is like a hash code for virus researchers as well. Less entropy you get, it would mean that it is likely encrypted or compressed code which could be potentially be a virus.

A standard binary would have a higher entropy than a compressed or encrypted one.

share improve this answer

Codingday
708 • 5 • 14

answered
Feb 4 '09 at 11:18

Interesting. I didn't know that. – **fluffels** Feb 4 '09 at 14:08

I think it must be the other way around. Entropy of encrypted-compressed code is higher than hand written code. e.g. a signal that is all 1's have 0 information, whereas the encrypted version would tend to have an equal probability of 1's or 0's just to hide the signal. in the case of compression (lossless), the information will be conserved (since... you should be able to

get back the original signal) but since the entropy is mean information, and we have fewer bits, entropy will be higher. Hope I am not missing sth. – **mehmet.ali.anil** Feb 24 '14 at 9:21

well.. regarding compression, a code that goes like FF00FF00FF00 can be compressed as 101010, or 2A, which has a higher entropy. – **mehmet.ali.anil** Feb 24 '14 at 9:27

add a comment

0

Entropy has many meanings typically in Computer Science. It depends on the context. In security entropy means how much randomality you place, for instance when you generate a private key many applications ask you to move the mouse around to generate entropy. This generates entropy by taking the "human" element of randomality and adds it to the hashing process of generating the key.

Now there is also a defnition for software engineering of entropy. This definition represents out of date code, or code that has had many developers writing it. Typically used in reference to when it is near time to refactor your software project. "The code for this project

has an enourmous amount of entropy because many of the individuals who maintained it are not on the project currently".

Here is a third example usage that I remembered too. In the topic of simulated annealing (as far as computer science is concerned), entropy is described as how much decay has happened during the evaluation of the algorithm.

I guess to answer your question though, there is not a concrete definition of the word 'entropy' except for the ones that you can find in a dictionary. How computer science tends to apply that term depends on the context of the term being used and what it is being applied to.

share improve this answer

jwendl
812 • 7 • 12

answered
Feb 7 '09 at 17:41

0

It's easy to make a big deal out of entropy. To my mind it is a pretty simple and useful concept.

Basically it quantifies what, on average, you will learn from an event, like flipping a coin, taking a branch instruction, or indexing an array.

Like a comparison operation in the middle of a search algorithm has a certain probability P of taking one branch, and 1-P of taking the other.

Suppose P is 1/2, as it is in a binary search. Then if you take that branch, you know 1 bit more than you did before, because log(2/1), base 2, is 1. On the other hand, if you take the other branch you also learn 1 bit.

To get the average amount of information you will learn, multiply what you learn on the first branch times the probability you take that branch, plus what you learn on the second branch times the probability of that branch.

1/2 times 1 bit, plus 1/2 times 1 bit, is 1/2 bit plus 1/2 bit, or total 1 bit of entropy. That's what you can expect to learn on average from that decision.

On the other hand, suppose you are doing linear search in a table of 1024 entries.

On the first == test, the probability of YES is 1/1024, so the entropy of YES at that decision

is

```
1/1024 times log(1024/1)
```

or 1/1024 * 10 = about 1/100 bit.

So if the answer is YES, you learn 10 bits, but the chance of that is about 1 in a thousand.

On the other hand, NO is much more likely. It's entropy is

```
1023/1024 * log(1024/1023)
```

or roughly 1 times roughly zero = about zero.

Add the two together, and on average you will learn about 1/100 of a bit on that decision.

That's why linear search is slow. The entropy (how much you can expect to learn) at each decision is too small, since you're going to have to learn 10 bits to find the entry in the table.

share improve this answer

Mike Dunlavey
32.7k • 7 • 65 • 103

answered
Mar 21 '09 at 1:09

0

Entropy in computer science commonly refers to how random a string of bits is. The following question is about making that precise:

How do I compute the approximate entropy of a bit string?

share improve this answer

dreeves
12.8k • 33 • 120 • 204

answered
Jun 5 '10 at 4:40

0

In simpler words, Entropy defines randomness. It’s more like how unpredictable something is. In more technical words, “In computing, entropy is the randomness collected by an operating system or application for use in cryptography or other uses that require random data. This randomness is often collected from hardware sources, either pre-existing ones such as mouse movements or specially provided randomness generators.” as defined by wikipedia.

One can now easily conclude the meaning of entropy in respect to a file as the measurement of the how much disordered the bytes are in a file. There are various units used for defining entropy like nat, shannon or hartley. Well, most common unit used is Shannon. The range of values a file’s entropy must come in as per Shannon’s algorithm is 0 to 8. So, when the entropy value is zero, one can say the outcome is certain. On contrary, when the entropy value is 8, the outcome is most unpredictable it could be. The formula given by Shannon to measure randomness in outcome of events is:

```
Entropy = ∑ pi log(1/pi)
```

where *i* is the event with probability *pi*.

This equation will always result in between 0 to 8.

For more information, go through the link: https://www.talentcookie.com/2016/02/file-

entropy-in-malware-analysis/

share improve this answer

Tarang Arora
124 • 3

answered
Feb 5 '16 at 15:03

tripleee
62.2k • 8 • 70 • 121

edited
Feb 5 '16 at 15:29

You assume 8-bit bytes without explaining it at all. The arbitrary range 0 to 8 makes no sense unless you realize it's 1 per bit. Still I hardly think this clarifies the matter. – **tripleee** Feb 5 '16 at 15:31

add a comment

-1

I've heard people misuse the thermodynamic definitions of entropy w.r.t CS.

E.g. Entropy is definitely increasing in this system.

When what they mean is this code is getting worse and worse!

share improve this answer

Fortyrunner
9,916 • 3 • 23 • 49

answered
Feb 4 '09 at 7:50

Your Answer

Body

Add picture

Log in

OR

Name

Email

By posting your answer, you agree to the privacy policy and terms of service.

Post Your Answer