Draft and Experiment Area

In [ ]:

```
The link to this report is
http://... .pdf
```

1. First impression

    - What is my chosen paper to read?

        > **"A Mathematical Theory of Communication" by Claude Shannon (1948)**

    - What type of the main contribution the paper has made?

        > **It quantified the fundamental limits on information transfer over noisy channels (channel capacity and error correction), and on the removal of information redundancy through data compression (source coding)**

    - *Before* reading the main body of the paper, write down your first impression obtained from its abstract and short introduction.

        > **It sets out to extend the work of Nyquist and Hartley to include channel noise and the reduced entropy in the transmission of English text, considering discrete, continuous and mixed channels. It sounds like it's going to get fairly mathematically complicated.**

    - Why does the paper attract you, such as, How it surprised you? Why do you think it addresses an important topic that will be helpful in your future study of machine learning?

        > **It's a paper whose time had come, due to the rise in importance of reliable electronic communication that came about after two world ars and Shannon's own work in cryptography during WWII, and the increased popularity of the telegraph alongside the beginnings of the computer age. I don't yet know what its relationship is to machine learning in particular.**

2. Read the paper abstract and introduction, list here all the notions that you don't know the precise meaning. If you think you have completed your list, compare the list with people around you who have chosen the same or a similar paper.

    > **PCM, PPM, entropy, Markoff process, ergodic, transducer**

3. (During the next 7 days) Re-consider the central problem of the paper.

Review Report on "A Mathematical Theory of Communication"

Introduction

In 1948 Dr Claude Shannon, a mathematician and electrical engineer, published a groundbreaking paper over two editions of the Bell System Technical Journal while employed at Bell Telephone Laboratories. "A Mathematical Theory of Communication" (Shannon 1948) set out to extend Nyquist's (1924) and Hartley's (1928) work on a general theory of communication by including mathematical theories of the effect of noise on the practical limits of channel capacity, as well as the limits and effectiveness of both error correction and data compression.

Content

In this paper, Shannon set out to address these five questions:

1. How to measure the *quantity of information?*
2. How to measure *channel capacity?*
3. What are the characteristics of efficient information *coding,* and what is the maximum *rate* at which an efficiently coded message can be transmitted over a channel?
4. How does channel *noise* affect message accuracy, how can its effect be minimised, and to what extent can it be eliminated?
5. How are these answers different when the signal is *continous* rather than discrete?

The answer to the first of these questions was to declare the **bit** as the fundamental unit of information, communicating to the receiver the resolution of the uncertainty of a binary possibility and therefore a reduction in a message's entropy.

To address the second question of channel capacity, the paper proposed theoretical limits to the information transfer rate for a nominated channel experiencing any specified noise level (the **Shannon limit** or **Shannon capacity**). This places a limit on the efficiency of error-correcting methods to counter the effects of noise and recover from consequent data corruption.

The paper addressed the third question (that of reducing unwanted message redundancy through coding and data compression) by showing that a lossless compression scheme cannot compress messages, on average, to have more than one bit of information per bit of message, but that any value less than one bit of information per bit of message can be attained by employing a suitable **coding scheme**. In addition, no lossless compression scheme can shorten **all** messages, with unlikely messages potentially made longer when coded. The limit of compression is the entropy of the message source, and compression **beyond** the channel capacity (beyond entropy) must result in information being lost. This means that, theoretically, it is possible to transmit information almost error-free at any rate below a limiting rate. Additionally, Shannon's prior work on cryptography during WWII told him that human communication consists of a mix of randomness and statistical language dependencies, with English text (which has fairly low entropy) containing approximately 50% redundancy, which is able to be exploited through data compression schemes.

For the fourth question, the paper showed how error-correcting methods could improve message accuracy over a noisy channel. Additionally, for any given degree of noise, discrete data can be communicated through the channel with an arbitrarily small error rate, **up to the channel capacity** (a computable maximum rate). Transmission at a rate beyond the channel capacity must inevitably result in the loss of information.

Shannon addressed the fifth question by showing that, although more difficult mathematically, the theory for a continuously varying message source is **not essentially different** from that for a discrete source. Although the paper derives upper and lower limits for channel capacity with a continuous source, it does not contain a single explicit formula for it in this more complicated case.

Innovation

While Nyquist (1924) concluded that the speed of "intelligence transmission" is proportional to the logarithm of the number of required symbols, and Hartley (1928) classed the receipt of information as the elimination of possibilities and the exclusion of other possible symbols, in this paper Shannon innovates by **formally integrating noise in the communication channel into his mathematical model of communication.**

The paper also contributes a **formal generalised model for communication,** consisting of the information source passing the message to a transmitter, which encodes a signal and sends it over a channel, where it is affected by noise. The receiver accepts the received signal (the signal influenced by the noise), which then decodes the message and passes it to the destination.

The adoption of the **bit** as the fundamental unit of information was also a key innovation of this paper, as it enabled the quantification of information in the detailed mathematical propositions that followed. Shannon's model of information was motivated by a desire to save time when sending messages over telegraph wires, but this led to the bit being adopted as a global unit of information and is now the foundation of our digital economy.

Finally, the paper shows how the redundancy present in every human language such as English defines the capacity of a channel carrying a message in such a language, proving mathematically that, **as channel noise increases, redundancy must also increase in order to communicate without error, which then decreases the information transfer rate.** Therefore, while it might seem efficient to remove this redundancy through coding and data compression in order to increase the transfer rate, this would only work on a noiseless channel, since the redundancy helps to combat the noise.

More generally, the innovative nature of this paper has been endorsed by many, with Horgan (1990) calling this paper "*the Magna Carta of the information age*", Lucky (1989) stating "*I know of no greater work of genius in the annals of technological thought*", and Solomon Golomb, a contemporary information theorist (quoted in Horgan 1992) stating "*It's like saying how much influence the inventor of the alphabet has had on literature*". Indeed, Weaver (1949) points out that "*This is a theory so general that one does not need to say what kinds of symbols are being considered - whether written letters or words, or musical notes, or spoken words, or symphonic music, or pictures. The theory is deep enough so that the relationships it reveals indiscriminately apply to all these and to other forms of communication*", and goes on to declare that "*the mathematical theory is exceedingly general in its scope, fundamental in the problems it treats, and of classic simplicity and power in the results it reaches*".

Technical quality

In general, the paper is of high technical quality, and Shannon provides suitable mathematical proofs or supports for all of his theoretical propositions. Additionally, each of his propositions has later been verified experimentally.

However, the paper's technical quality is lacking in at least three aspects. Firstly, although the paper includes a description of the theoretical maximum for the efficiency of error correction in the presence of noise, *Shannon gave only an outline of the proof for the discrete case.* A more rigorous mathematical proof would have to wait until Feinstein (1954).

Secondly, while the theorem considered both the case in which the information transmission rate is above the channel capacity and the case in which it is below it, *it does not address the rare situation in which rate and capacity are essentially equal.* This oversight is unfortunate and seems inexplicable.

Lastly, the paper did not take into account *the effects of significant transmission times over the channel on the correct interpretation of the message by the receiver.* We have all suffered the confusing effects of attempting to have a conversation with someone on the other side of the world, where the transmission delay causes speech collisions with the consequent loss of information and the need for repetition of missed phrases. Therefore, even with a noiseless channel, information can be lost solely due to transmission latency when two people are attempting a conversation.

Application and X-factor

As a universally acknowledged foundational paper that formed the basis for the entire field of Information Theory, the theoretical propositions put forward in this paper have been proven to be appropriate for this domain. However, particularly in the area of error-correction (channel coding) techniques, it took many years for these theoretical propositions to be put into reality, and to identify methods that the paper showed were possible. Like much of Einstein's theoretical work, this only serves to illustrate the groundbreaking character of Shannon's work in this paper.

There have been many application domains in which this research work has already been applied:

- lossless data compression in the field of data compression tools such as ZIP files;
- lossy data compression in the field of media file encoding techniques such as MP3 and JPEG files;
- error-correcting codes in the field of media encoding such as CDs and DVDs;
- channel coding in the field of digital data transmission over phone lines such as DSL;
- entropy and information gain calculation for the construction of decision trees using the ID3 and C4.5 algorithms in machine learning.

Even outside of the Information Technology field, this work has informed measures of entropy in communities, brain research in psychology, Hawking's theories about information leakage from Black Holes, and many other fields.

Given this pedigree, it is difficult to imagine new ways in which the foundational ideas presented in this paper could be applied in as-yet-untouched application domains. One such interesting area would be to *apply the ideas of information entropy from this paper to the dance language of honey bees.* It is well-established that the ability of the scout bee to quickly and effectively communicate the location of floral resources is key to a hive's viability. Therefore, measuring the effectiveness of the coding and the reduction in entropy produced by different components of the dance language may lead to a better understanding of social behaviour in bees. Another potential area is the *study of ant behaviours when communicating the location of a food source,* (which involves chemical signalling through pheromones, as well as touch, body language and sound), interpreted as a problem of information coding, transmission over a noisy channel, decoding and message interpretation.

In terms of possible further developments of Shannon's research work, NASA's *planned manned missions to Mars* will require significant research into optimal methods of data compression, error correction, and other aspects of attempting reliable communication over an extremely noisy channel for which it takes between 3 and 22 minutes (depending on the relative planet positions) for a signal from Earth to reach Mars or vice versa. Advanced error-correction techniques such as Reed-Solomon codes, low-density parity-check (LDPC) and turbo codes have taken us closer to the Shannon limit, but it is not yet known whether further progress can be made in this area. In addition, work done in this area will also benefit communications back on Earth, as we attempt to squeeze ever more information (often streaming standard-definition or high-definition video) over our already overstretched Internet infrastructure.

Presentation

The paper's overall structure is reasonable, starting with an overall Introduction followed by separate sections for discrete noiseless systems, discrete channels with noise, continuous channels and continuous source rates. However, as it was published in two separate sections, the now-complete paper has four appendices with no conclusion at the end of the first section, and another three appendices with no conclusion at the end of the second section. I found this confusing, obfuscating the overall flow of the paper and hindering understanding. In a paper of 55 pages, with heavy mathematical content, more discussion of the meaning of the results and a well-written conclusion would definitely aid comprehension.

When it was originally published, it was found to be difficult for less mathematically-inclined scholars to understand, as it appears to have been written (at least in part) for an audience of fellow Bell Labs engineers and mathematicians. This was evident to such a pronounced degree that Weaver (1949) was asked to interpret it using less formidable language and for a more general audience, and this was subsequently published along with Shannon's original paper as "The Mathematical Theory of Communication" (Shannon and Weaver 1949). Indeed, I found it difficult to follow the paper's argument, partly due to the sheer quantity of formulae and appendices and partly due to my lack of suitable engineering background to understand all of the mathematics. However, my shortcomings cannot detract from Shannon's achievement in the publication of this landmark paper.

References

Feinstein, A. 1954, 'A new basic theorem of information theory', *Transactions of the IRE Professional Group on Information Theory,* vol. 4, no. 4, pp. 2-22.

Hartley, R.V.L. 1928, 'Transmission of information', *The Bell System Technical Journal,* vol. 7, no. 3, pp. 535-63.

Horgan, J. 1990, 'Claude E. Shannon: Unicyclist, juggler, and father of information theory', *IEEE Information Theory Society Newsletter,* June 1990.

Horgan, J. 1992, 'Claude E. Shannon', *IEEE Spectrum,* April, 1992, pp. 72-5.

Lucky, R.W. 1989, *Silicon Dreams: Information, Man, and Machine,* St. Martin's Press, New York, p. 37

Nyquist, H. 1924, 'Certain factors affecting telegraph speed', *The Bell System Technical Journal,* vol. 3, no. 2, pp. 324-46.

Rogers, E.M. 1994, 'Claude Shannon's cryptography research during World War II and the mathematical theory of communication', *1994 Proceedings of IEEE International Carnahan Conference on Security Technology,* pp. 1-5.

Seising, R. 2009, '60 years "A Mathematical Theory of Communication" - Towards a "Fuzzy Information theory"', *2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference, IFSA-EUSFLAT 2009 - Proceedings,* pp. 1332-7.

Shannon, C.E. 1948, 'A mathematical theory of communication', *The Bell System Technical Journal,* vol. 27, no. 3, pp. 379-423, and no. 4, pp. 623-56.

Shannon, C.E. & Weaver, W. 1949, *The mathematical theory of communication,* Univ. of Chicago Press, Urbana.

Draft and Experiment Area

In [ ]:

1. First impression

   - What is my chosen paper to read?

     > **"A Mathematical Theory of Communication" by Claude Shannon (1948)**

   - What type of the main contribution the paper has made?

     > **It quantified the fundamental limits on information transfer over noisy channels (channel capacity and error correction), and on the removal of information redundancy through data compression (source coding)**

   - *Before* reading the main body of the paper, write down your first impression obtained from its abstract and short introduction.

     > **It sets out to extend the work of Nyquist and Hartley to include channel noise and the reduced entropy in the transmission of English text, considering discrete, continuous and mixed channels. It sounds like it's going to get fairly mathematically complicated.**

   - Why does the paper attract you, such as, How it surprised you? Why do you think it addresses an important topic that will be helpful in your future study of machine learning?

     > **It's a paper whose time had come, due to the rise in importance of reliable electronic communication that came about after two world ars and Shannon's own work in cryptography during WWII, and the increased popularity of the telegraph alongside the beginnings of the computer age. I don't yet know what its relationship is to machine learning in particular.**

2. Read the paper abstract and introduction, list here all the notions that you don't know the precise meaning. If you think you have completed your list, compare the list with people around you who have chosen the same or a similar paper.

   > **PCM, PPM, entropy, Markoff process, ergodic, transducer**

3. (During the next 7 days) Re-consider the central problem of the paper.

Review Report on "A Mathematical Theory of Communication"

Introduction

In 1948 Dr Claude Shannon, a mathematician and electrical engineer, published a groundbreaking paper over two editions of the Bell System Technical Journal while employed at Bell Telephone Laboratories. "A Mathematical Theory of Communication" (Shannon 1948) set out to extend Nyquist's (1924) and Hartley's (1928) work on a general theory of communication by including mathematical theories of the effect of noise on the practical limits of channel capacity, as well as the limits and effectiveness of both error correction and data compression.

Content

In this paper, Shannon set out to address these five questions:

1. How to measure the *quantity of information?*
2. How to measure *channel capacity?*
3. What are the characteristics of efficient information *coding,* and what is the maximum *rate* at which an efficiently coded message can be transmitted over a channel?
4. How does channel *noise* affect message accuracy, how can its effect be minimised, and to what extent can it be eliminated?
5. How are these answers different when the signal is *continous* rather than discrete?

The answer to the first of these questions was to declare the **bit** as the fundamental unit of information, communicating to the receiver the resolution of the uncertainty of a binary possibility and therefore a reduction in a message's entropy.

To address the second question of channel capacity, the paper proposed theoretical limits to the information transfer rate for a nominated channel experiencing any specified noise level (the **Shannon limit** or **Shannon capacity**). This places a limit on the efficiency of error-correcting methods to counter the effects of noise and recover from consequent data corruption.

The paper addressed the third question (that of reducing unwanted message redundancy through coding and data compression) by showing that a lossless compression scheme cannot compress messages, on average, to have more than one bit of information per bit of message, but that any value less than one bit of information per bit of message can be attained by employing a suitable **coding scheme**. In addition, no lossless compression scheme can shorten **all** messages, with unlikely messages potentially made longer when coded. The limit of compression is the entropy of the message source, and compression **beyond** the channel capacity (beyond entropy) must result in information being lost. This means that, theoretically, it is possible to transmit information almost error-free at any rate below a limiting rate. Additionally, Shannon's prior work on cryptography during WWII told him that human communication consists of a mix of randomness and statistical language dependencies, with English text (which has fairly low entropy) containing approximately 50% redundancy, which is able to be exploited through data compression schemes.

For the fourth question, the paper showed how error-correcting methods could improve message accuracy over a noisy channel. Additionally, for any given degree of noise, discrete data can be communicated through the channel with an arbitrarily small error rate, **up to the channel capacity** (a computable maximum rate). Transmission at a rate beyond the channel capacity must inevitably result in the loss of information.

Shannon addressed the fifth question by showing that, although more difficult mathematically, the theory for a continuously varying message source is **not essentially different** from that for a discrete source. Although the paper derives upper and lower limits for channel capacity with a continuous source, it does not contain a single explicit formula for it in this more complicated case.

Innovation

While Nyquist (1924) concluded that the speed of "intelligence transmission" is proportional to the logarithm of the number of required symbols, and Hartley (1928) classed the receipt of information as the elimination of possibilities and the exclusion of other possible symbols, in this paper Shannon innovates by **formally integrating noise in the communication channel into his mathematical model of communication.**

The paper also contributes a **formal generalised model for communication,** consisting of the information source passing the message to a transmitter, which encodes a signal and sends it over a channel, where it is affected by noise. The receiver accepts the received signal (the signal influenced by the noise), which then decodes the message and passes it to the destination.

The adoption of the **bit** as the fundamental unit of information was also a key innovation of this paper, as it enabled the quantification of information in the detailed mathematical propositions that followed. Shannon's model of information was motivated by a desire to save time when sending messages over telegraph wires, but this led to the bit being adopted as a global unit of information and is now the foundation of our digital economy.

Finally, the paper shows how the redundancy present in every human language such as English defines the capacity of a channel carrying a message in such a language, proving mathematically that, **as channel noise increases, redundancy must also increase in order to communicate without error, which then decreases the information transfer rate.** Therefore, while it might seem efficient to remove this redundancy through coding and data compression in order to increase the transfer rate, this would only work on a noiseless channel, since the redundancy helps to combat the noise.

More generally, the innovative nature of this paper has been endorsed by many, with Horgan (1990) calling this paper "*the Magna Carta of the information age*", Lucky (1989) stating "*I know of no greater work of genius in the annals of technological thought*", and Solomon Golomb, a contemporary information theorist (quoted in Horgan 1992) stating "*It's like saying how much influence the inventor of the alphabet has had on literature*". Indeed, Weaver (1949) points out that "*This is a theory so general that one does not need to say what kinds of symbols are being considered - whether written letters or words, or musical notes, or spoken words, or symphonic music, or pictures. The theory is deep enough so that the relationships it reveals indiscriminately apply to all these and to other forms of communication*", and goes on to declare that "*the mathematical theory is exceedingly general in its scope, fundamental in the problems it treats, and of classic simplicity and power in the results it reaches*".

Technical quality

In general, the paper is of high technical quality, and Shannon provides suitable mathematical proofs or supports for all of his theoretical propositions. Additionally, each of his propositions has later been verified experimentally.

However, the paper's technical quality is lacking in at least three aspects. Firstly, although the paper includes a description of the theoretical maximum for the efficiency of error correction in the presence of noise, *Shannon gave only an outline of the proof for the discrete case.* A more rigorous mathematical proof would have to wait until Feinstein (1954).

Secondly, while the theorem considered both the case in which the information transmission rate is above the channel capacity and the case in which it is below it, *it does not address the rare situation in which rate and capacity are essentially equal.* This oversight is unfortunate and seems inexplicable.

Lastly, the paper did not take into account *the effects of significant transmission times over the channel on the correct interpretation of the message by the receiver.* We have all suffered the confusing effects of attempting to have a conversation with someone on the other side of the world, where the transmission delay causes speech collisions with the consequent loss of information and the need for repetition of missed phrases. Therefore, even with a noiseless channel, information can be lost solely due to transmission latency when two people are attempting a conversation.

Application and X-factor

As a universally acknowledged foundational paper that formed the basis for the entire field of Information Theory, the theoretical propositions put forward in this paper have been proven to be appropriate for this domain. However, particularly in the area of error-correction (channel coding) techniques, it took many years for these theoretical propositions to be put into reality, and to identify methods that the paper showed were possible. Like much of Einstein's theoretical work, this only serves to illustrate the groundbreaking character of Shannon's work in this paper.

There have been many application domains in which this research work has already been applied:

- lossless data compression in the field of data compression tools such as ZIP files;
- lossy data compression in the field of media file encoding techniques such as MP3 and JPEG files;
- error-correcting codes in the field of media encoding such as CDs and DVDs;
- channel coding in the field of digital data transmission over phone lines such as DSL;
- entropy and information gain calculation for the construction of decision trees using the ID3 and C4.5 algorithms in machine learning.

Even outside of the Information Technology field, this work has informed measures of entropy in communities, brain research in psychology, Hawking's theories about information leakage from Black Holes, and many other fields.

Given this pedigree, it is difficult to imagine new ways in which the foundational ideas presented in this paper could be applied in as-yet-untouched application domains. One such interesting area would be to *apply the ideas of information entropy from this paper to the dance language of honey bees.* It is well-established that the ability of the scout bee to quickly and effectively communicate the location of floral resources is key to a hive's viability. Therefore, measuring the effectiveness of the coding and the reduction in entropy produced by different components of the dance language may lead to a better understanding of social behaviour in bees. Another potential area is the *study of ant behaviours when communicating the location of a food source,* (which involves chemical signalling through pheromones, as well as touch, body language and sound), interpreted as a problem of information coding, transmission over a noisy channel, decoding and message interpretation.

In terms of possible further developments of Shannon's research work, NASA's *planned manned missions to Mars* will require significant research into optimal methods of data compression, error correction, and other aspects of attempting reliable communication over an extremely noisy channel for which it takes between 3 and 22 minutes (depending on the relative planet positions) for a signal from Earth to reach Mars or vice versa. Advanced error-correction techniques such as Reed-Solomon codes, low-density parity-check (LDPC) and turbo codes have taken us closer to the Shannon limit, but it is not yet known whether further progress can be made in this area. In addition, work done in this area will also benefit communications back on Earth, as we attempt to squeeze ever more information (often streaming standard-definition or high-definition video) over our already overstretched Internet infrastructure.

Presentation

The paper's overall structure is reasonable, starting with an overall Introduction followed by separate sections for discrete noiseless systems, discrete channels with noise, continuous channels and continuous source rates. However, as it was published in two separate sections, the now-complete paper has four appendices with no conclusion at the end of the first section, and another three appendices with no conclusion at the end of the second section. I found this confusing, obfuscating the overall flow of the paper and hindering understanding. In a paper of 55 pages, with heavy mathematical content, more discussion of the meaning of the results and a well-written conclusion would definitely aid comprehension.

When it was originally published, it was found to be difficult for less mathematically-inclined scholars to understand, as it appears to have been written (at least in part) for an audience of fellow Bell Labs engineers and mathematicians. This was evident to such a pronounced degree that Weaver (1949) was asked to interpret it using less formidable language and for a more general audience, and this was subsequently published along with Shannon's original paper as "The Mathematical Theory of Communication" (Shannon and Weaver 1949). Indeed, I found it difficult to follow the paper's argument, partly due to the sheer quantity of formulae and appendices and partly due to my lack of suitable engineering background to understand all of the mathematics. However, my shortcomings cannot detract from Shannon's achievement in the publication of this landmark paper.

References

Feinstein, A. 1954, 'A new basic theorem of information theory', *Transactions of the IRE Professional Group on Information Theory,* vol. 4, no. 4, pp. 2-22.

Hartley, R.V.L. 1928, 'Transmission of information', *The Bell System Technical Journal,* vol. 7, no. 3, pp. 535-63.

Horgan, J. 1990, 'Claude E. Shannon: Unicyclist, juggler, and father of information theory', *IEEE Information Theory Society Newsletter,* June 1990.

Horgan, J. 1992, 'Claude E. Shannon', *IEEE Spectrum,* April, 1992, pp. 72-5.

Lucky, R.W. 1989, *Silicon Dreams: Information, Man, and Machine,* St. Martin's Press, New York, p. 37

Nyquist, H. 1924, 'Certain factors affecting telegraph speed', *The Bell System Technical Journal,* vol. 3, no. 2, pp. 324-46.

Rogers, E.M. 1994, 'Claude Shannon's cryptography research during World War II and the mathematical theory of communication', *1994 Proceedings of IEEE International Carnahan Conference on Security Technology,* pp. 1-5.

Seising, R. 2009, '60 years "A Mathematical Theory of Communication" - Towards a "Fuzzy Information theory"', *2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference, IFSA-EUSFLAT 2009 - Proceedings,* pp. 1332-7.

Shannon, C.E. 1948, 'A mathematical theory of communication', *The Bell System Technical Journal,* vol. 27, no. 3, pp. 379-423, and no. 4, pp. 623-56.

Shannon, C.E. & Weaver, W. 1949, *The mathematical theory of communication,* Univ. of Chicago Press, Urbana.