부호화 지문 분석 - 의식의 흐름 기법 사용

아 기술 지문 싫은뎅...

디지털 통신 시스템은 송신기, 채널, 수신기로 구성되며, ⓐ전송할 데이터를 빠르고 정확하게 전달하기 위해 부호화 과정을 거쳐 전송한다. 영상, 문자 등인 데이터는 ⓑ기호 집합에 있는 기호들의 조합이다. 예를 들어 기호 집합 {a, b, c, d, e, f}에서 기호들을 조합한 add, cab, beef 등이 데이터이다.

질문할 거리가 없네 그냥 받아들여야 하나
알았어 기억하고 있을게

정보량은 어떤 기호가 발생했다는 것을 알았을 때 얻는 정보의 크기이다. 어떤 기호 집합에서 특정 기호의 발생 확률이 높으면 그 기호의 정보량은 적고, 발생 확률이 낮으면 그 기호의 정보량은 많다. 기호 집합의 평균 정보량*을 기호 집합의 엔트로피라고 하는데 모든 기호들이 동일한 발생 확률을 가질 때 그 기호 집합의 엔트로피는 최댓값을 갖는다.

* 평균 정보량: 각 기호의 발생 확률과 정보량을 서로 곱하여 모두 더한 것.

뭔데!
왜 발생 확률이 높아야 정보량이 적은데?
그리고 왜 동일한 발생 확률을 가져야 엔트로피,
즉 평균 정보량이 커지는데?
아직 모르겠다

송신기에서는 소스 부호화, 채널 부호화, 선 부호화를 거쳐 기호를 ⓒ부호로 변환한다. 소스 부호화는 데이터를 압축하기 위해 기호를 0과 1로 이루어진 부호로 변환하는 과정이다. 어떤 기호가 110과 같은 부호로 변환되었을 때 0 또는 1을 비트라고 하며 이 부호의 비트 수는 3이다. 이때 기호 집합의 엔트로피는 기호 집합에 있는 기호를 부호로 표현하는 데 필요한 평균 비트 수의 최솟값이다.

어... 어... 어?
기호 집합의 엔트로피가 또 나오네
근데 바로 윗문단에서는 평균 정보량이라며
왜 달라진 건데!
어 잠시만 달라진게 아니라면
평균 정보량 = 필요한 평균 비트 수의 최솟값인거?
아 그러면 평균 정보량이 작아지면
필요한 평균 비트 수가 작아지는 거구나

전송된 부호를 수신기에서 원래의 기호로 ⓓ복원하려면 부호들의 평균 비트 수가 기호 집합의 엔트로피보다 크거나 같아야 한다.

위 내용을 참고하면
비트 수가 많을 수록 담을 수 있는 정보 량이 많아지니까
부호들의 평균 비트수가 기호집합의 평균정보량(엔트로피)
보다 커야 기호 집합의 정보를 다 담겠구나

기호 집합을 엔트로피에 최대한 가까운 평균 비트 수를 갖는 부호들로 변환하는 것을 엔트로피 부호화라 한다. 그중 하나인 '허프만 부호화'에서는 발생 확률이 높은 기호에는 비트 수가 적은 부호를, 발생 확률이 낮은 기호에는 비트 수가 많은 부호를 할당한다.

왜 가장 가까운 평균 비트수를 갖는 부호들로 변환해야 함?
그리고 허프만 부호화 외에 그렇게 해야함?
아 그리고
허프만 부호화 굳이 왜 알려줬니?
근데 허프만 부호화의 이유는 알겠네
그래야 사용하는 비트 수가 가장 적어지니까
자주 쓰는 기호는 간단한 부호로 나타내야 편하지
어라 그러면 허프만 부호화는 사용하는 비트 수를
최대한 줄이는 과정에서 생성하는 거네
즉 비트수를 최대한 엔트로피에 가깝게
부호화 하는 거니까 허프만 부호화가 엔트로피
부호화구나!

채널 부호화는 오류를 검출하고 정정하기 위하여 부호에 잉여 정보를 추가하는 과정이다. 송신기에서 부호를 전송하면 채널의 잡음으로 인해 오류가 발생하는데 이 문제를 해결하기 위해 잉여 정보를 덧붙여 전송한다.

잉여 정보 붙여서 어쩌게? 오류가 그걸로 해결됨?
아까는 비트수 줄이려고 노력하더만
왜 정보를 늘리는거야?

채널 부호화 중 하나인 '삼중 반복 부호화'는 0과 1을 각각 000과 111로 부호화한다. 이때 수신기에서는 수신한 부호에 0이 과반수인 경우에는 0으로 판단하고, 1이 과반수인 경우에는 1로 판단한다. 즉 수신기에서 수신된 부호가 000, 001, 010, 100 중 하나라면 0으로 판단하고, 그 이외에는 1로 판단한다. 이렇게 하면 000을 전송했을 때 하나의 비트에서 오류가 생겨 001을 수신해도 0으로 판단하므로 오류는 정정된다.

아 그래
잉여 정보를 붙여서 보내면 오류가 생겨도 정정되네

채널 부호화를 하기 전 부호의 비트 수를, 채널 부호화를 한 후 부호의 비트 수로 나눈 것을 부호율이라 한다. 삼중 반복 부호화의 부호율은 약 0.33이다.

부호율이 그래서 네 글에서 무슨 의미를 갖는데?
내가 한번 의미 부여를 해볼까
부호율이 작다는 건... 추가되는 잉여 정보가
많아진다는 거고 그러면 오류 정정에 좋되겠네
대신 정보량이 많아지겠다.