5.1 정보량

우선 한 사건 (event) 에서 기대되는 정보량(quantity of information, I) 을 어떻게 나타낼 수 있는지를 생각해 보자. 이 정보량을 수량화하기 위하여 확률 (P) 과 정보량 (I) 을 Gawron(2010) 에 따라 다음의 두 확률 조건의 관점에서 살펴보자.

1. 중요성 (significance): 어떤 사건이 일어날 가능성이 작으면 작을수록, 그 사건은 더 많은 정보를 지닌다.
   $P(x_1) > P(x_2) \Rightarrow I(x_1) < I(x_2)$

2. 가법성(additivity): 만일 x_1, x_2 가 독립적인 사건이라면 다음을 만족해야 한다.
   $I(x_1 x_2) = I(x_1) + I(x_2)$

중요성 조건은 어떤 사건의 확률이 높을수록 이 사건으로 알려지는 정보량은 적어짐을 나타낸다. 따라서 확률값을 역으로 취하여 중요성에 따른 정보량을 나타낼 수 있다.

$$I(x) = 1/P(x) \tag{5.1}$$

예제 5.1 확률과 정보량의 관계는 다음과 같이 예시될 수 있다.

$P(x_1) = \frac{1}{2}$ 이면 $I(x_1) = \frac{1}{1/2} = 2$
$P(x_2) = \frac{1}{4}$ 이면 $I(x_2) = \frac{1}{1/4} = 4$

따라서 중요성의 조건이 만족된다. 만일 두 사건이 서로 독립적이라면,

$$P(x_1 x_2) = P(x_1) \times P(x_2) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$$
$$I(x_1 x_2) = \frac{1}{P(x_1 x_2)} = 8$$

이다. 그러나 가법성에 따라

$$I(x_1x_2) = I(x_1) + I(x_2) = 2 + 4 = 6$$

이다. 따라서 가법성의 조건이 충족되지 못한다.

두 독립 사건의 확률값은 곱으로 이루어지지만, 두 사건의 결합된 정보 내용은 더해져야만 한다. 정보의 가법성을 위해서 곱이 아닌 더하기가 필요하다. 따라서 이와 유사한 기능을 하는 log를 도입하게 된다.[1]

$$\log(xy) = \log x + \log y \tag{5.2}$$

즉 확률을 역으로 하여 중요성 기준을 만족시킬 수 있고, log를 사용하여 가법성의 조건을 만족시킬 수 있다. 이제 이 둘을 결합하면 어떤 확률 변수 x가 지니는 정보량은 다음과 같이 계산될 수 있다.

$$I(x) = \log \frac{1}{P(x)} = -\log P(x) \tag{5.3}$$

예제 5.2 예제 5.1을 로그를 사용하여 계산해 보자.

$P(x_1) = \frac{1}{2}$ 이면 $I(x_1) = -\log(\frac{1}{2}) = \log 2 = 1$
$P(x_2) = \frac{1}{4}$ 이면 $I(x_2) = -\log(\frac{1}{4}) = \log 4 = 2$

$P(x_1) > P(x_2) \Rightarrow I(x_1) < I(x_2)$ 인 중요성의 조건을 만족한다. 또,

$$P(x_1x_2) = P(x_1)P(x_2) = \tfrac{1}{8} \Rightarrow I(x_1x_2) = -\log(\tfrac{1}{8}) = 3$$
$$I(x_1x_2) = I(x_1) + I(x_2) = 1 + 2 = 3$$

으로 가법성의 조건도 만족하게 된다.

[1] 정보 이론에서 사용되는 로그는 대개 밑이 2인 로그를 의미한다. 이 장에서 사용되는 로그는 특별한 명시가 없는 한 밑이 2인 로그다.

이렇게 계산된 정보량은 밑이 2 인 로그로 계산되기 때문에 그 단위는 비트 (bit) 가 된다. 따라서 확률변수 X 에 대한 정보량 (I) 은 5.3 과 같이 확률값을 역으로 취한 로그값으로 구할 수 있다. 이 로그로 표시되는 값은 그 사건에 의해 생성되는 놀라움의 정도 (amount of surprise) 라고 할 수 있다. 확률값이 작으면 작을수록 이 사건에 의해 나타나는 놀라움의 정도는 커지게 된다. 확률값이 크면 클수록 우리는 그 사건에 의해 나타날 결과를 더 많이 예측할 수 있기 때문에, 그 결과에 대한 놀라움의 정도는 적어진다.

5.2 엔트로피

확률 변수 X 의 표본 공간에서 나타나는 모든 사건들의 정보량의 평균적인 기댓값(expectation) 을 엔트로피라 하고 H 로 표시한다. 엔트로피는 자연언어처리와 음성인식에 있어서 중요한 역할을 한다. 예를 들어 엔트로피는 언어 처리를 위해 설정된 모델이 어느 정도 그 언어에 적합한 정보를 지니고 있으며, 또 실제 언어에 얼마나 잘 부합하는지 등을 측정할 수 있는 방법을 제공한다. 일반적으로 엔트피로의 계산은 어떤 확률 변수가 지니는 평균적인 불확실성 (uncertainty) 을 측정하여 이루어지며 다음의 수식으로 표시된다.

$$H(p) = H(X) = -\sum_{x \in X} P(x) \log_2 P(x) \tag{5.4}$$

엔트로피는 이 수식에서 볼 수 있듯이, 한 확률 변수에서 나타나는 개별적인 사건의 확률과 그 정보량을 곱하고 다시 이를 모두 더하여 얻어진 기댓값이다. 또한 밑이 2 인 로그를 사용하기 때문에 비트 (bit) 에 관한 정보도 나타낸다.

예제 5.3 어떠한 알려진 정보도 없이, 1 번부터 8 번까지 8 마리 말이 참가하는 경마에서 매 경기마다 우승하는 말의 번호를 확률변수 X 로 하는