Axioms of Probability

Let (Ω, A) be a measurable Space of events.

(i) for a Event $E \in A$.

$$P(E) \geq 0 \quad \text{(Non-negativity)}$$

(ii) for two event E_1 $E_2 \in A$.

$$E_1 \subseteq E_2$$

$$P(E_1) \leq P(E_2) \geq 0$$

(iii) $P(\Omega) = 1$

(iv) $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

(v) for E_1 and E_1'

$$P(E_1) = 1 - P(E_1')$$

Conditional Probability.

$$P(A/B)$$

(i) A and B are independent.

$$P(A/B) = P(A)$$

(ii) A and B are dependent

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

(iii) A and B are disjoint / mutually exclusive

$$P(A/B) = 0$$

(iv) Product rule.

$$P(A,B) = P(A/B) \cdot P(A)$$
$$= P(B/A) \cdot P(B)$$

Generalizing

$$P(A_1 A_2 \ldots A_n) = P(A_n / A_1 \ldots A_{n-1}) \cdot P(A_1 \ldots A_{n-1})$$

Baye's Rule

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

$P(A)$ = Prior Probability of hypothesis
$P(B)$ = Evidence
$P(B/A)$ = Likelihood
$P(A/B)$ = Posterior

Odd of A/B

$$O(A/B) = \frac{P(A/B)}{P(\neg A/B)}$$

$$= \frac{P(A/B)}{1 - P(A/B)}$$

Probability Ranking Principle

Let

$D = \{d_1, d_2 \ldots d_n\}$ $\quad |D| = n$

q is a fixed query.

As, Per Probability ranking Principle (PRP) for a given 'q' we can rank documents as decreasing order of probability of relevance.

$$\forall_{d_i=1}^{n} \text{ rank drecuny } P[R=1/d,q]$$

$P[R=1/\vec{x},\vec{q}]$

$P[R=1/\langle t_1 \ldots t_n \rangle, \vec{q}]$

This ranking of documents are same if we use probability of odd

that is

$$O[R/\vec{x},\vec{q}] = \frac{P[R=1/\vec{x},\vec{q}]}{P[R=0/\vec{x},\vec{q}]}$$

$$O\left[R/\vec{x},\vec{q}\right] = \frac{P\left[R=1/\vec{q}\right]\cdot P\left[\vec{x}/R=1,q\right]}{P(\vec{x}/\vec{q})}$$

$$\div \quad \frac{P\left[R=0/\vec{q}\right]\cdot P\left[\vec{x}/R=0,q\right]}{P(\vec{x}/\vec{q})}$$

Here, $P(\vec{x}/\vec{q})$ and $P\left[R=1/\vec{q}\right]/P\left[R=0/\vec{q}\right]$ is constant as per Assumption

Hence

$$O\left[R/\vec{x},\vec{q}\right] = \frac{P\left[\vec{x}/R=1,\vec{q}\right]}{P\left[\vec{x}/R=0,\vec{q}\right]}$$

$$= \bigwedge_{t=1}^{M} \frac{P\left[\vec{x}_t/R=1,\vec{q}\right]}{P\left[\vec{x}_t/R=0,\vec{q}\right]}\cdot O(R/\vec{q})$$

Let

$$p_t = P(x_t=1/R=1,\vec{q})$$

$$u_t = P(x_t=1/R=0,\vec{q})$$

So,

	$R=1$	$R=0$
$x_t=1$	p_t	u_t
$x_t'=0$	$1-p_t$	$1-u_t$

$$O[R/\vec{x},\vec{q}] = \prod_{t=x_t=q_t=1} \frac{p_t}{u_t} \cdot \prod_{t\,x_t=0\,q_t=1} \frac{1-p_t}{1-u_t}$$

Now

$$\prod_{t\,x_t=0\,q_t=1} \frac{1-p_t}{1-u_t}$$ is a constant

$$O[R/\vec{x},\vec{q}] = \prod_{t=x_t=q_t=1} \frac{p_t(1-u_t)}{u_t(1-p_t)} \cdot \prod_{t=q_t=1} \frac{1-p_t}{1-u_t}.$$

This odd $[R=1/\vec{x},\vec{q}]$ can be used as a Retrieval Status value

$$RSV = \prod_{t=x_t=q_t=1} \frac{p_t(1-u_t)}{u_t(1-p_t)}$$

$$= \prod_{t=x_t=q_t=1} \frac{p_t/(1-p_t)}{u_t/(1-u_t)}$$

we can use logarithm to transform this into summation.

Estimating P_t and u_t.

- Let N be the # of Doc. in the corpus.
- R be the # relevant documents in our sample
- n_t be the number of documents containing term t.
- r_t be the # of relevant Doc. that contain t

$$P_t = \frac{r_t}{R} \qquad u_t = \frac{n_t - r_t}{N - R}$$

Lidstone smoothing ($\lambda = 0.5$)

$$P_t = \frac{r_t + 0.5}{R + 1} \qquad u_t = \frac{n_t - r_t + 0.5}{N - R + 1}$$

Example