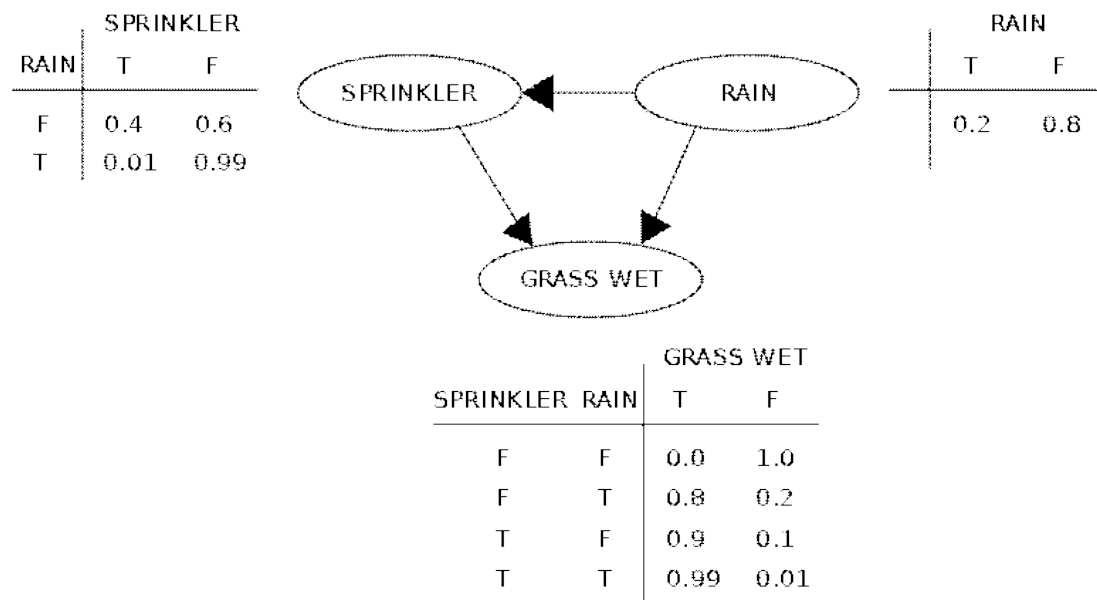


# Information Retrieval Modelling:

## Bayesian network models



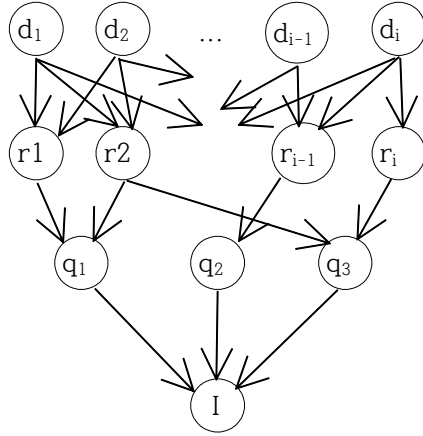


Figure 6: A simplified inference network

도 6에 있는 모든 노드들은  $\{0,1\}$ 의 값을 갖는 이진 무작위 변수를 나타낸다. 이론적으로 이 모델의 작동방법을 알기 위하여, 노드들의 하위집합, 예를 들어  $r_2$ ,  $q_1$ ,  $q_3$  와  $I$ 을 살펴보는 것이 도움이 될 것이다. 그리고 지금은 나머지 노드들을 무시하자. 확률의 연쇄법칙(chain rule)에 의해, 노드  $r_2$ ,  $q_1$ ,  $q_3$  와  $I$ 의 결합 확률은 다음과 같다:

$$P(r_2, q_1, q_3, I) = P(r_2)P(q_1|r_2)P(q_3|r_2, q_1)P(I|r_2, q_1, q_3) \quad (\text{공식 11})$$

원호의 방향은 무작위 변수들의 간의 의존적 관계를 나타낸다. event “information need is fulfilled” ( $I = 1$ )은 2 가지의 잠재적 이유(causes)를 갖는다: 쿼리 노드  $q_1$ 이 참이거나 쿼리 노드  $q_3$  가 참이다(기억할 것은 우리는  $q_2$ 를 무시한다는 것이다). 반대로 이 두 개의 쿼리 노드는 표현 노드  $r_2$  에 의존하고 있다. 그러므로 이 모델은 다음과 같은 조건적 독립성 가정을 만든다:

$$P(r_2, q_1, q_3, I) = P(r_2)P(q_1|r_2)P(q_3|r_2)P(I|q_1, q_3) \quad (\text{공식 12})$$

오른쪽에, 3번째 확률척도는 단순화되었는데, 왜냐하면  $q_1$ 과  $q_3$ 가 자신들의 parent  $r_2$  에 독립적이기 때문이다. 마지막 부분인  $P(I|q_1, q_3)$ 는 단순화되었는데, 그 이유는  $I$ 가 그것의 parent  $q_1$ 과  $q_3$  일 경우,  $r_2$  와 독립적이기 때문이다.

이 네트워크에서 직선적 사용은 쿼리 노드의 수가 너무 많다면 실질적이지 못하다. 하나의 노드에 특정화되어야 하는 확률의 수는 parents의 수와 더불어 지수적으로 늘어난다. 예를 들어,  $n$ 개의 쿼리 노드를 갖고 있는 네트워크는 정보요구 노드용으로  $P(I|q_1, q_2, \dots, q_n)$ 의  $2^{n+1}$ 인 잠재적 값을 필요로 한다. 이런 이유로 인하여, 모든 네트워크 레이어들은 어떤 형태의 approximations를 필요로 한다. Metzler와 Croft는 approximation 에 대해 다음과 같

이 설명하였다: “다큐 레이어용으로, 이들이 가정한 것은 단지 한 개의 싱글 다큐만이 동시에 (at a time) 관찰되며, 모든 싱글 다큐용으로는 그것의 다큐 레이어를 무시한 분리된 (separate) 네트워크가 구축된다”는 것이다.“ 모든 네트워크의 표현 레이어용으로, 표현 노드의 확률(which effectively are priors now, because the document layer is ignored)은 어떤 검색모델에 의해 평가될 수 있다. 주목할 것은 표현 노드는 대체로 단일 용어이지만 어구일 수도 있다는 것이다. 최종적으로, 쿼리 노드와 정보요구 노드는 소위 believe operators에 의해 정의되는 표준 확률분산에 의해 근접된다(approximated). 이 연산자들은 고정된 방식으로 표현 노드와 기타 쿼리 노드의 확률 값을 결합시킨다. 만일  $P(q_1|r_2)$ 와  $P(q_3|r_2)$ 의 값이  $p_1$ 과  $p_2$ 에 의해 주어진다면,  $P(I|r_2)$ 의 계산은 *and*, *or*, *sum* 그리고 *wsum*과 같은 연산자를 사용하여 얻게 될 것이다.

$$\begin{aligned} P_{\text{and}}(I|r_2) &= p_1 \cdot p_2 \\ P_{\text{or}}(I|r_2) &= 1 - ((1-p_1)(1-p_2)) \\ P_{\text{sum}}(I|r_2) &= (p_1 + p_2)/2 \\ P_{\text{wsum}}(I|r_2) &= w_1p_1 + w_2p_2 \end{aligned} \quad (\text{공식 13})$$

각 연산자마다 공식 13에서처럼 계산될 수 있는, 예를 들어  $P(I|q_1, q_3)$ 의 정의가 존재하는 소위 link matrices가 이러한 연산자용으로 존재한다는 것을 알 수 있다. 따라서 비록 연산자가 속한 link matrix가 거대해질 수 있다 하더라도, 이것은 실제로 존재하지 않으며 그 결과는 linear time 으로 계산할 수 있다. 그렇더라도 누군가 주장할 수 있는 것은 만일 아직까지 그러한 시도가 “Bayesian network model”이라고 하는 것에 의존한다면, 각 네트워크 레이어에 대한 approximations가 의심스럽다는 것이다.

▪ Bayes 정리: 베이즈 확률계산(영어: Bayes' theorem)는 두 확률 변수의 사전 확률과 사후 확률 사이의 관계를 나타내는 정리다

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

- ▲  $\Pr(A|B)$ -도서관: B의 값이 주어진 경우에 대한 A의 사후 확률.
- ▲  $\Pr(B|A)$ : A가 주어졌을 때 B의 조건부 확률.
- ▲  $\Pr(A)$ : A의 사전확률으로, 아직 사건 B에 관한 어떠한 정보도 알지 못한다는 것을 의미한다.
- ▲  $\Pr(B)$ : B의 사전확률이며, 정규화 상수의 역할을 한다.

<예>

눈앞의 사람에 대한 사전확률이 주어졌을 경우에, 철수의 사후확률을 구하는 공식

$$P(\text{철수}|\text{눈앞의사람}) = \frac{P(\text{눈앞의사람}|\text{철수})P(\text{철수})}{P(\text{눈앞의사람})}$$

▲ point

P(눈앞의사람)도 사실 계산이 가능한 상수값인데, 만일 가능한 사람이 철수, 영희, 순희 3 사람이라고 한다면,

$$P(\text{눈앞의사람}) =$$

$$P(\text{눈앞의사람}|\text{철수})P(\text{철수}) + P(\text{눈앞의사람}|\text{영희})P(\text{영희}) + P(\text{눈앞의사람}|\text{순희})P(\text{순희})$$

$$= \left(\frac{1}{3} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{1}{3}\right) = \frac{3}{9} = \frac{1}{3}$$

- 끝 -