베이즈 망 (Bayesian Network)

베이즈 망 (베이지안 네트워크)는 확률론적 그래픽 모델로, 변수들 간의 조건부

의존성을 나타내는 방향성 비순환 그래프 (Direct Acyclic Graph, DAG)이다.

이는 복잡한 시스템의 이해와 예측이 가능하며, 불확실한 상황에서도 효과적인 결정을

내릴 수 있다.

베이즈 망의 구성 요소에는 노드와 엣지가 있다.

노드는 변수들을 나타낸다. 각 노드는 하나의 랜덤 변수를 의미하며, 이 변수는 다양한

상태를 가질 수 있다. 예를 들어, 질병 진단 시스템에서는 각 노드가 특정 질병 또는

증상을 나타낼 수 있다.

* 변수 종류: 노드는 이산 변수 또는 연속 변수일 수 있다. 이산 변수는

유한한 상태를 가지며, 연속 변수는 특정 범위 내에서 모든 실수 값을 가질

수 있다.

* 확률 분포: 각 노드는 그 자신과 관련된 확률 분포를 가진다. 부모 노드가

없는 노드는 사전 확률을 가지며, 부모 노드가 있는 노드는 조건부 확률

분포 (conditional probability distribution, CPD)를 가진다.

엣지는 노드들 간의 관계를 나타내며, 방향성이 존재한다. 엣지의 방향은 인과 관계를

의미하며, 이는 하나의 노드가 다른 노드에 조건부로 영향을 미친다는 것을 의미한다.

* 조건부 의존성: 부모 노드와 자식 노드 간의 조건부 의존성을 나타낸다.

자식 노드는 부모 노드의 상태에 의존하여 확률 분포를 가진다.

* 전이 관계: 엣지의 방향은 한 변수의 상태 변화가 다른 변수의 상태에

영향을 미치는 방식을 나타낸다.

이러한 베이즈 망의 핵심 요소는 조건부 확률 분포다. 각 노드는 그 부모 노드들이

주어졌을 때의 조건부 확률을 가지고 있다. 예시로, 노드 A와 그 부모 노드 B, C가

있다면, P(A|B, C)라는 조건부 확률로 표현된다. 전체 네트워크의 확률 분포는 모든

노드의 조건부 확률의 곱으로 표현된다.

이 수식에서의 Xi는 네트워크의 각 노드를 나타내고, Parents(Xi)는 노드 Xi의 부모

노드들을 의미한다.

베이즈 망은 베이즈 정리를 기반으로 작동한다. 그 기반이 되는 베이즈 정리는 다음과 같다.



이를 통해 새로운 증거가 주어졌을 때, 이의 사후 확률을 계산할 수 있다. 예시로, 질병

진단 문제에서 특정한 증상이 나타났을 때, 해당 증상을 고려한 질병의 확률을 계산할

수 있다.

다음은 베이즈 망에서의 두 가지 주요 추론 방법이다.

* 정확한 추론: 모든 가능한 경우를 고려하려 정확한 확률 값을 계산한다. 이는

주로 작은 네트워크에서 사용된다.

* 근사 추론: 큰 네트워크에서는 정확한 추론이 비현실적이므로 근사적인

방법을 주로 사용한다. 대표적인 방법으로는 샘플링 기법 (Monte Carlo

Method)과 변분 추론 (Variational Inference) 등이 있다.

베이즈 망의 학습 방법은 크게 두 가지로 나눌 수 있다.

* 구조 학습: 데이터에서 네트워크의 구조를 학습한다. 이는 노드들 간의

관계를 추정하는 과정이다. 대표적인 알고리즘으로는 K2 알고리즘과

Hill-Climbing 등이 있다.

* 매개변수 학습: 주어진 구조에서 조건부 확률 분포를 학습한다. 완전한

데이터가 주어졌을 때에는 최대우도추정법 (Maximum Likelihood

Estimation, MLE)을 사용하고, 결측 데이터가 있는 경우에는 기대 최대화

알고리즘 (Expectation-Maximization, EM)을 사용한다.

베이즈 망은 다양한 분야에서 사용되고, 응용된다.

* 의료 진단: 증상을 기반으로 질병을 진단하는 데에 사용된다.
* 유전자 네트워크 분석: 유전자 간의 상호작용을 이해하고 예측하는 데에

활용된다.

* 자연어 처리: 문맥을 고려한 단어의 의미를 추론하는 데에 사용된다.
* 재무 모델링: 주식 가격 예측 및 리스크 관리에 활용된다.