

探索的モデリング (第 1 回)

井庭崇

iba@sfc.keio.ac.jp

1 ベイジアンネットワークと隠れマルコフモデル

1.1 ベイジアンネットワーク

「ベイジアンネットワーク」は、定性的な因果関係のネットワークと定量的な確率が組み合わさったモデルで、事象の結果から原因を推定できるという特徴があります。このモデルは、機械学習、データマイニング、ユーザーモデリング、ロボティクス、ゲノム解析、医療診断、故障診断、意思決定支援システムなど、幅広い分野に応用されています。

1.2 隠れマルコフモデル

「隠れマルコフモデル」は、事象の発生過程を状態遷移ネットワークを用いて統計的に予測するモデルで、観測可能なデータからその現象の背後にある隠れた構造を推定することができます。このモデルは、最も成功をおさめた音声認識の他に、ゲノム解析や顔の認識研究などに応用されています。

2 結果から原因を推定ということ：ベイズ推論

今、「発熱」の症状がある患者が目の前にいる。あなたは医者であり、その発熱の原因を診断したいと思っている。ここでは、発熱には、「風邪」、「コレラ」、「赤痢」という3つの原因が考えられるとして、どれが原因かを推定したい。

最もありえそうな原因を選ぶ 最も大きい確率の原因を選ぶということ。そこで、「結果 B が起こったときに原因が A である確率」を知りたい。

つまり、次の条件付き確率を求めたい。

条件付き確率 $P(\text{原因 } A | \text{結果 } B)$

$P(\text{風邪} | \text{発熱})$

$$P(\text{コレラ} | \text{発熱})$$

$$P(\text{赤痢} | \text{発熱})$$

しかし、確率 $P(\text{原因 } A | \text{結果 } B)$ を直接求めるのは難しい。

ふつうは、「原因 A が起こった場合に、結果が B になる確率」 $P(\text{結果 } B | \text{原因 } A)$ が分かっている場合がほとんど。

例えば、発熱の場合では、

$$P(\text{発熱} | \text{風邪}) \quad \text{風邪のときに発熱する確率}$$

$$P(\text{発熱} | \text{コレラ}) \quad \text{コレラのときに発熱する確率}$$

$$P(\text{発熱} | \text{赤痢}) \quad \text{赤痢のときに発熱する確率}$$

は、過去のデータからそれなりに計算できる。

そこで、工夫が必要。「ベイズの定理」を使う。

「原因 A が起こった場合に結果が B になる確率」 $P(\text{結果 } B | \text{原因 } A)$ から、「結果 B が起こったときに原因が A である確率」 $P(\text{原因 } A | \text{結果 } B)$ を計算する。

最終的に、例えば、

$$P(\text{風邪} | \text{発熱}) = 0.8$$

$$P(\text{コレラ} | \text{発熱}) = 0.7$$

$$P(\text{赤痢} | \text{発熱}) = 0.9$$

のように確率が計算でき、それから、原因の診断をする。

参考文献

- 月本 洋, 松本 一教, 『やさしい 確率・情報・データマイニング』, 森北出版, 2004