1995年度日本オペレーションズ・リサーチ学会
秋季研究発表会

ベイズの定理を構成したニューラルネットワークによる指数分布の信頼度分布表現

篠原康秀　尾崎俊治
広島大学工学部

1. はじめに

近年，ニューラルネットワークを因果関係がはっきり分かっていない問題に応用した報告が多数ある．これらは，そのような複雑な問題に対しニューラルネットワークが有効であることを示しているが，その結果に対し理論的な補償を示した報告は少ない [1].

一方，信頼性の分野では，アイテムが高信頼性になったため信頼度を求めることが難しくなってきている．そこで，エキスパートのもつ知識を用いて信頼度関数を表現する方法が報告されている [2]．しかしながら客観性という観点からみれば，エキスパートのもつ知識は人間の主観であるために得られた結果について疑問が残るといえる．

したがって，エキスパートによって決める信頼度関数のパラメータをニューラルネットワークによって求めることにより客観性を確立し，更に，ニューラルネットワークの構造にベイズの定理を組み込むことによって，アイテムを支配するパラメータをネットワークが獲得できるようにする．つまり，指数分布にしたがうデータを学習する過程で，ネットワークが故障率 λ を獲得するため，結果について理論的な補償をすることができる．

本研究では，指数分布にしたがうデータに対し，ベイズの定理を構造としてもつニューラルネットワークを用いて信頼度分布表現を行うことに関して報告する．また，ニューラルネットワークの学習終了後にネットワークの各パラメータについて検討する．

2. 論理的背景

2.1 指数分布の故障率 λ とベイズの定理を用いた尤度

故障率 λ に対するデータ d_i の条件付き確率は，ベイズの定理によって次式で与えられる．

$$P(d_i|\lambda)=\frac{P(d_i)P(\lambda|d_i)}{\sum_{i=1}^{n}P(d_i)P(\lambda|d_i)} \tag{1}$$

ここで，n はデータの個数である．

一方，データ d $(=d_1,d_2,\cdots,d_n)$ についての尤度 $L(d|\lambda)$ は，

$$L(d|\lambda)=\prod_{i=1}^{n}P(d_i|\lambda) \tag{2}$$

で与えられるから，(1) 式より (2) 式は，

$$L(d|\lambda)=\frac{\prod_{i=1}^{n}P(d_i)P(\lambda|d_i)}{\sum_{i=1}^{n}P(d_i)P(\lambda|d_i)} \tag{3}$$

である．

2.2 三層ニューラルネットワーク

三層ニューラルネットワークは，入力層，隠れ層，出力層から構成される．各層は任意の数のユニットからなり，入力層のユニットは隠れ層のユニットに結合係数を介してユニットの出力値を伝える．各ユニットは，シグモイド関数によりその出力値を出力する．また，生理学的根拠からシグモイド関数にはオフセット値が含まれている．

一般に，シグモイド関数は次式で与えられる．

$$f(x)=\frac{1}{1+\exp(-(x+\theta)/T)} \tag{4}$$

ここで，x はシグモイド関数に与えられる入力値．θ はオフセット値．T はネットワークの温度である．

3. ベイズの定理を構造として持つニューラルネットワーク

ここでは，前節で述べた (3) 式をニューラルネットワークにより構成する．まず，故障データと同数のユニットからなる入力層では，シグモイド関数よりデータ d に対し確率 $P(\lambda|d_i)$ が計算され，入力層出力とする．次に，$\hat{\lambda}$から算出した結合係数 $P(d_i)$ を乗じて入力層出力は隠れ層に入力される．そして，隠れ層の各ユニット出力は，(1) 式によって計算される．最後に，隠れ層からの入力値 $P(d_i|\lambda)$ の積を計算して出力層入力とし，出力層から尤度 $L(d|\lambda)$ を出力する．

データから$\hat{\lambda}$を計算し，教師信号としてネットワークの学習に用いる．学習終了後，結合係数 $P(d_i)$ より信頼度分布を求める．

3.1 学習則

3.節で述べたニューラルネットワークの学習は，結合係数 $P(d_i)$，ユニットのオフセット値θ，ネットワークの温度 Tを変更することによって行うことにする．

教師信号$\hat{\lambda}$と出力層 $L(d|\lambda)$ との誤差δは，次式で与えられる．

$$\delta=(\hat{\lambda}-L(d|\lambda))^2/2 \tag{5}$$

1. 結合係数 $P(d_i)$ の修正量 $\Delta P(d_i)$

したがって，結合係数 $P(d_i)$ の微小変化に対する誤差$\partial\delta/\partial P(d_i)$ は，

$$\frac{\partial\delta}{\partial P(d_i)} \;=\; -(\hat{\lambda}-L(d|\lambda))\prod_{j\neq i}^{n}P(d_j|\lambda)\,P(\lambda|d_i)$$

$$\times \quad (\sum_{j=1}^{n} P(\lambda|d_j)P(d_j) - P(\lambda|d_i)P(d_i))$$

$$\times \quad (\sum_{j=1}^{n} P(\lambda|d_j)P(d_j))^{-2} \qquad (6)$$

で与えられる．したがって修正量$\Delta P(d_i)$は，上式から次式により求められる．

$$\Delta P(d_i) = -\alpha \frac{\partial \delta}{\partial P(d_i)} \qquad (7)$$

ここで，αは正の定数である．上式を用いて結合係数 $P(d_i)$ を修正することにより，教師信号との誤差を減らすことができる．

2. オフセット値θ_i の修正量 $\Delta\theta_i$
上式と同様に，オフセット値θの微小変化に対する誤差$\partial\delta/\partial\theta_i$ は次式で与えられる．

$$\begin{aligned}\frac{\partial \delta}{\partial \theta_i} &= -(\hat{\lambda} - L(d|\lambda)) \prod_{j \neq i}^{n} P(d_j|\lambda)\, P(d_i) \\ &\times \quad (\sum_{j=1}^{n} P(\lambda|d_j)P(d_j) - P(\lambda|d_j)P(d_j)) \\ &\times \quad (\sum_{j=1}^{n} P(\lambda|d_j)P(d_j))^{-2} \\ &\times \quad (1 - f(d_i))f(d_i)/T_i \qquad (8)\end{aligned}$$

したがって，オフセット値の修正量$\Delta\theta_i$は上式を用いて次式で与えられる．

$$\Delta\theta_i = -\beta \frac{\partial \delta}{\partial \theta_i} \qquad (9)$$

ここで，βは正の任意定数である．

3. ネットワークの温度T_i の修正量 ΔT_i 同様に温度Tの微小変化に対する誤差$\partial\delta/\partial T_i$は，

$$\begin{aligned}\frac{\partial \delta}{\partial T_i} &= -(\hat{\lambda} - L(d|\lambda)) \prod_{j \neq i}^{n} P(d_j|\lambda)\, P(d_i) \\ &\times \quad (\sum_{j=1}^{n} P(\lambda|d_j)P(d_j) - P(\lambda|d_j)P(d_j)) \\ &\times \quad (\sum_{j=1}^{n} P(\lambda|d_j)P(d_j))^{-2} \\ &\times \quad -(1 - f(d_i))f(d_i)/T_i^2 \qquad (10)\end{aligned}$$

ここで，温度Tが増加すると，$\partial\delta/\partial T$は減少するので，修正係数$\gamma$は負になる．したがって，温度の修正量は，

$$\Delta T_i = -\gamma \frac{\partial \delta}{\partial T_i} \qquad (11)$$

よって，(7)，(9)，(11) 式を用いて学習を行う．

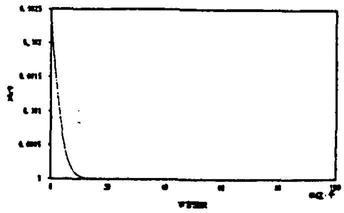

Figure 1: 学習によって更新された出力値

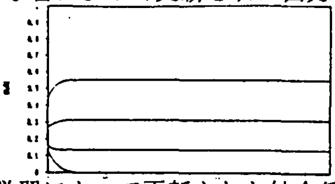

Figure 2: 学習によって更新された結合係数

4. 数値例

ここでは，データ **d** として，110, 300, 500, 600 を与えシミュレーションを行う．学習を繰り返すことによりネットワークが尤度を獲得する様子を図 1に示す．

また，結合係数 $P(d_i)$ の変化の様子を図 2に示す．
学習により更新された結合係数から，信頼度分布を求め図 3に示す．

5. 考察

学習回数が非常に大きいのは，出力層において積を計算しているために誤差の修正量$\Delta P(d_i), \Delta\theta_i, \Delta T_i$ が極めて小さくなるからである．しかしながら，学習回数が進むにつれて尤度が出力されることが示された．また，$\hat{\lambda}$から求めた結合係数がネットワークの学習によって更新されることも示した．更に，学習終了時点での結合係数値から求めた信頼度分布は，$\hat{\lambda}$を用いて算出した指数分布と比べて問題がないと思われる．

6. まとめ

本研究で示したニューラルネットワークは尤度を求めるものであるが，故障率 λの元で条件付き確率を学習するため，このネットワークから求めた信頼度分布について理論的な補償がされているといえる．従って，本研究で提案した信頼度分布の表現法は，有効な方法であるといえる．

References

[1] 辻敏夫，森大一郎，伊藤宏司．統計的構造を組み込んだニューラルネットワークによる EMG 動作識別法．電学論 C,112 巻 8 号, 平成 4 年．

[2] N.D.Singpurwalla. An Interactive PC-Based Procedure for Reliability Assessment Incorporating Expert Opinion & Survival Data. JASA, Vol.83, No.401, pp.43-51, 1988.

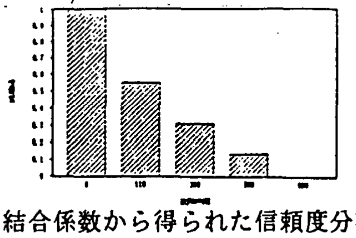

Figure 3: 結合係数から得られた信頼度分布