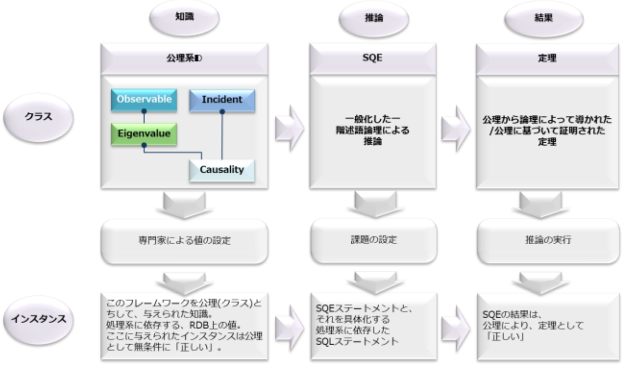
F-002

深層学習における公理的知識モデル

中村　正治
Masaharu NAKAMURA

1．エクスパートシステムにおける公理論的知識

2017 年の FiT 論文にて、筆者は、フレームワークを公理系とする Expert System の構築法とそこから導かれる結論が公理として正しさを保証されるメカニズムを示した。



この整然とした構造により、与えられた知識を公理系の中で定理として導くことにより、正しさが保証されることと、論理式を SQE という形でリフォームして、現実の問題に適用できるようにした。

2．機械学習の論理構造へのアプローチ

一方、機械学習に代表される、ニューラルネットベースのシステムは、その構造に対する俯瞰的なヴューが今一つはっきりしていなかった。

本論では、ニューラルネットの幾何学的構造をまず明らかにして、ニューラルネットを成立させている数学的条件を定式化する。

次に、ニューラルネットの演算が正解に近づくことができる、ということがまさに、ニューラルネット・システム成立の公理となっていることを明らかにする。

ただし、本校においては、残念ながら表題で目指した、ニューラルネットにおける知識を公理的に定義する、という結論には至らなかった。これは次のテーマとしたい。

3．単層ニューラルネットの幾何学的構造

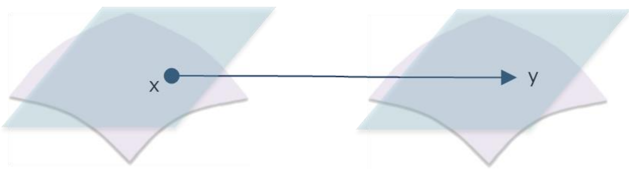
ここで、まず、単層のニューラルネットの幾何構造を調べ、それを逆に定義として定式化する。

EM を m 次元のリーマン多様体(m-fold)とし、事象多様体と名付ける。また、RM を同じく n 次元リーマン多様体(n-fold)とし、結果多様体と名付ける。
x∊ EM, y∊ RM としたとき、x, y 上の接空間のベクトルをそれぞれ
$\mathbf{u}=(u^1,u^2\cdots u^m)$, $\mathbf{v}=(v^1,v^2\cdots v^n)$と表す。

以降、すべて、この x, y の近傍を扱うが、煩雑なのでいちいち近傍とは断らない。

この時
$v^a=\sum_b p^a{}_b u^b$ とする p が存在するならば、これを評価関数と呼ぶ。
(なお、添字の上下についての説明は省略するが、本論では、これ以降、上下で同じ添字がある場合はΣ記号を省略する(所謂アインシュタイン表記)ので、上記変換式は単に $v^a=p^a{}_b u^b$ と表記される。)

y 上の接空間のベクトルを $\mathbf{vc}=(vc^1,vc^2\cdots vc^n)$と表したとき、$|\mathbf{vc}-\mathbf{v}|$を最小にすることを p の最適化と呼ぶ。



次に、D を共変微分として、x, y の近傍にそれぞれ x+dx,y+dy をとったとき、D を共変微分として、

評価関数 q $v^a+Dv^a=q1^a{}_b(u^b+Du^b)$ について、q=p が成り立つとき、(EM,RM,p)をニューラルネットと呼ぶ。

多くの場合、m>n であり、p は正方行列ではなく、また写像としては Homeomorphic ではないことに留意が必要である。

さてこの時、CE 上の接続を G, CR 上の接続を H として、

$D_{y|a}v^b=\partial_{y|a}v^b+H_{ac}{}^b v^c$
$\quad=\partial_{y|a}p^b{}_c u^c+H_{ac}{}^b p^c{}_d u^b$

そもそも、

$D_{y|a}v^b dy^a=D_{x|a}u^b dx^a$ なので、

p が x の近傍で一定とする条件により、

$(\partial_{y|a}+H_{ac}{}^b)\,p^c{}_d u^b=D_{x|c}u^b\,\partial x^c/\partial y^a$
$\quad=\partial_{y|a}x^b\cdot(\partial_{x|a}+G_{ac}{}^b)u^c$

これが、評価関数の満たすべき方程式となる。

4．単層ニューラルネットの幾何学的構造

次にこれを多層ニューラルネットに拡張する。

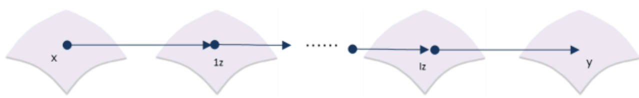
EM を m 次元のリーマン多様体(m-fold)とし、事象多様体と名付ける。また、RM を同じく n 次元リーマン多様体(n-fold)とし、結果多様体と名付ける。

1M, 2M,・・・IM というリーマン多様体を仮定して、中間層多様体と命名し、それぞれの次元を ie とする。

x∈ EM, y∈ RM としたとき、

X, Y 上の接空間のベクトルをそれぞれ $\mathbf{u}=(u^1,u^2\cdots u^m)$, $\mathbf{v}=(v^1,v^2\cdots v^n)$、

1z∈ 1M, 2z∈ 2M, ・・・iZ∈ iM,・・・(1≦i≦I) として、iZ 上の接空間のベクトルを $\mathbf{iw}=(iw^1,iw^2\cdots iw^{ie})$とする。



この時　j=i-1 として、
i=1 のとき　　$iw^a = ip^a_b u^b$
I>i≧2 のとき　$iw^a = ip^a_b jw^b$
i=I のとき　　$v^a = ip^a_b jw^b$

これらにより、
$$v^a = ip^a_b jw^b$$
$$= ip^a_b jp^b_c jw^c$$
$$= \cdots$$
$$= 1p^a_b\, 2p^b_c \ldots Ip^s_t u^t$$
$P^a_t = 1p^a_b\, 2p^b_c \ldots Ip^s_t$ とおけば、
$v^a = P^a_b u^b$　として、単層の場合と同じように記述できる。

評価関数は、
$$(\partial_{y|a} + H_{ac}{}^b)\, P^c{}_d u^b = D_{x|c} u^b\, \partial x^c / \partial y^a$$
$$= \partial_{y|a} x^b \cdot (\partial_{x|a} + G_{ac}{}^b) u^c$$

これが、評価関数の満たすべき方程式となる。
ここにおいて、注目すべきことは、この方程式が、EM と RM の接続のみを含んでいて、iM の接続を含まないことである。したがって、中間層多様体は接続項のない、平らなユークリッド多様体として通常の微分ですべて事足りることになる。

5．ニューラルネットの論理構造

以上の議論から、ニューラルネットは、幾何学的な構造からは、多層、単層と同等に扱ってもよさそうである。このときの、論理構造を検討する。
今までの議論から、
x∊EM, y∊RM, $\mathbf{u}$ を x 上のベクトル、$\mathbf{v}$ を y 上のベクトル、P を評価関数とすると、$\mathbf{v}=P\mathbf{u}$ と書ける。
このとき、P の定義により、x の近傍 x’上のベクトル $\mathbf{u'}$, y の近傍 y’上のベクトル $\mathbf{v'}$ においても, $\mathbf{v'}=P\mathbf{u'}$が成り立つ。
いま、y 上の「正解」を $\mathbf{vc}$ としたとき、P’を P の最適化として $\mathbf{vc}=P'\mathbf{u}$ が存在する。
この δP = P’ - P を 0 にする P が存在する、ということが a priori に仮定されているのが、ニューラルネットの論理構造の出発点となる。
すなわち、x∊EM 上の接空間を $ET_{|x}$,y∊RM 上の接空間を $RT_{|y}$ としたとき、P を評価関数、P’を P の最適化, δP =P’-P とすると、

$$\forall \mathbf{u} \in ET_{|x} \exists \mathbf{v} \in RT_{|y} | \mathbf{v}=P\mathbf{u} \quad \rightarrow$$
$$(\exists \mathbf{vc} \in RT_{|y} \exists P' \mid \mathbf{vc}=P'\mathbf{u}) \wedge (\delta P=0)$$

これがニューラルネットによる推論システムが成り立つための公理だといえる。

6．結論

単層・多層をあわせて、ニューラルネットの幾何学的構造と、そこから導かれる、ニューラルネットの公理を導いた。今後、このシステムにおける、「知識」の構造を引き続き探求する予定である。