高温相反学習と複雑なスピンモデル

湘南工科大学　　野倉一男

連想記憶モデルにおける学習では、本来学習したものとは異なる擬状態が系に生じることが知られている。これを取り除く機構として反学習という考えが提案され、ホップフィールドモデルに対するシュミレーションでは確かに連想記憶の能力が改善されることが示されてきた。生物学的には、反学習はレム睡眠中に起こり、擬状態は夢に対応すると仮定されている。モデルの変化の解析的な研究は、擬状態そのものの反学習という形式ではかなり困難と思われる。しかし、これを高温相のスピン状態の反学習に置き換えると、相互作用変化はスピン相関関数で表され、高温展開を用いてもとの相互作用で表すことができるようになる。

この定式化によれば、初期条件をホップフィールドモデルにとると反学習の結果、より能力が高い pseudo-inverse model が現れることが示される。このモデルはかなり以前に提案されていたが、モデルの構成に人工的な側面があり生物学的な意味は疑われていた。しかし反学習の理論はこのような神経回路が自然界に存在しうることを示唆している。

反学習の考えは大変一般的で、どのようなスピンモデルにも適用できる。初期条件をＳＫモデルにとる場合は、相互作用に相関が生ずるということのほかに、次のような生物学的な事実からも興味がある。すなわち、ヒトにおいて出生から数年間は神経回路が急速に、おそらくランダムに発達する時期であるが、この時期のレム睡眠の時間は成人にくらべて大変長い。従ってこのとき神経回路のランダムな形成と反学習とが同時におこると考えられる。ＳＫモデルを初期条件にしたときは反学習によるモデルの変化を簡単な形で表すのは難しいが、発展方程式よりアンチホップフィールドモデル（ホップフィールドモデルにおいて相互作用の符号を逆にしたもの）が変化したモデルの性質を反映していると期待される。このモデルのエネルギー関数は反学習との関係のほかに、簡単な意味をもつ最適化問題のコスト関数と見ることもできる。

反学習によって現れる pseudo-inverse model やアンチホップフィールドモデルはＳＫモデルとは大変異なったスピングラスモデルである。特にこれらのモデルでは相互作用行列の最大固有値がマクロに縮退している。このような特徴をもつスピンモデルとしては、数年前に相互作用行列の固有値が±1 をとるようなスピンモデルが調べられ、動的な相転移の存在が示された。この相転移点はＳＫモデルのようにレプリカ対称な解が現れる温度としては求まらない。我々の数値計算によればアンチホップフィールドモデルについても同様に動的な相転移が存在する。これは相互作用行列の最大固有値の縮退によってエネルギー関数もかなり縮退しており、スピン空間にスピンダイナミクスの固定点が大変多く存在することの反映と考えられる。これを超短期記憶の起源と考えてよいかどうかは今後の課題であるが、反学習によってランダムスピンモデルに質的な変化がおこるということはレム睡眠による神経回路の発達を理解する上で興味ある結果である。

Paramagnetic unlearning and Complex spin models

Kazuo NOKURA
Shonan Institute of Technology, Fujisawa, Japan

We discuss the spin models which are created by paramagnetic unlearning in neural network models. The spin models which arise by this evolution are quite different from the Hopfield model and usual spin glass models such as the SK model. One of the important features is that the largest eigen values of the interaction matrix degenerate macroscopically, i.e. the number of them is of order of the system size. As an example of these models, we present the studies of the anti-Hopfield model, the Hopfield model with an opposite interactional sign.

The idea of unlearning in neural networks was proposed several years ago as a mechanism to remove spurious states during REM sleep[1]. Numerical studies have given affirmative results to the neural network model which was defined by the energy function

$$H = -\frac{1}{2}\sum_{ij} J_{ij}S_iS_j. \tag{1}$$

This idea is very interesting not only in the study of neural-network evolution but also in the study of complex spin models. In the framework of statistical mechanics, it is natural and convenient to unlearn paramagnetic configurations instead of spurious states[2]. This corresponds to assuming the paramagnetic temperature in the formulation of the Boltzmann machine. The resulting interactional changes are expressed by the high temperature spin correlation functions, which can be studied by high temperature expansion. In this way, we arrive at the evolution equation of the form

$$J_{ij}' = cJ_{ij} - \epsilon\sum_{k} J_{ik}J_{kj}. \tag{2}$$

The studies of equation (2) revealed that, when initial model is assumed to be the Hopfield model, the model evolves into the pseudo-inverse(PI) model[2]. This result is very important for several reasons. Firstly, since the evolution rule is local, the PI model becomes biologically relevant by the process of unlearning. Secondly, as an associative memory, the PI model is better than the Hopfield model not only because it has greater capacity but also because it can memorize correlated patterns, which are confused in the Hopfield model. These points support the biological speculations about the function of dream sleep.

We are also interested in what happens when the initial model is the spin glass model. We can see that frustrations among interactions are changed by (2). In addition to this academic interest, unlearning in random neural networks can play important role in the development of neural networks since some biological observations revealed that newborns, whose neural networks presumably develop randomly, spend much longer time in REM sleep that adults. Here, instead of studying the evolution, we discuss the interactions without the first terms in (2), assuming uncorrelated randomness for J_{ij}. This model is very similar to the Hopfield model with an opposite interactional sign, i.e. the anti-Hopfield(AH) model. In addition to the relation to unlearning, we can show that the AH model has a simple meaning as an optimization problem.

To see the nature of the AH model, it is illuminating to study eigen value distributions of the interaction matrices. The AH model with $\alpha \equiv P/N < 1$ (N:the system size, P:the number of patterns) has strongly degenerated largest eigen values. This situation is very similar to the PI model and the random orthogonal(RO) model, which was recently studied in a different context[3]. The reason for this phenomenon is that the evolution equation(2), when it is written in terms of eigen values of interaction matrix, has only one stable fixed point to which positive eigen values tend.

The replica study of the AH model revealed that it has replica symmetry instability at $T_s = -1 + \sqrt{\alpha}$, which implies that there are no spin glass phase transitions for $\alpha < 1$. However in computer simulations, we found a sharp change in the temperature dependence of the internal energies at finite temperature just like the RO model.

Following the same idea as developed in [3], we found that, for $\alpha < 1$, the transition point is well identified by studying the one-step replica symmetry breaking(RSB) solutions with the marginality condition, which is supposed to characterize dynamical phase transitions. The Edwards-Anderson order parameters are very close to 1 even near the transition point. On the other hand for $\alpha >> 1$, the results of simulation agree with the usual RSB solution just like the SK model. These aspects are consistent with the properties of eigen value distributions of interactions. The existance of dynamical phase transitions implies that there are many fixed points of spin dynamics in rather high energies, which are separated by tall energy barriers.

What is the meaning of these results in the study of neural networks? In the Hopfield model, it works to remove the spurious states which are created by learning. In the spin glass model, it removes the low energy states which are strongly sample-dependent. Thus, in the context of neural networks, unlearning will suppress this property and increase the chances to learn about environment. In both cases, the energy landscapes of spin glass states become more rugged as unlearning proceeds. This will cause the increase of remnant properties, which presumably work as very short-term memories. This situation seems quite consistent with very long REM sleep of newborns, since they know nothing about what to learn and need to adape to any information outside. This point is an open question and I hope that our studies help to clarify the function of dream sleep.

[1]F.Crick and G Mitchison, Nature 304(1983)111

[2]K.Nokura, J.Phys A29(1996)3871, Phys.Rev. E54(1996)5571

[3]E.Marinari, G.Parisi, and F.Ritort, J.Phys. A27(1994)7647