2010 年 7 月 2 日
情報システム工学科 塩貝亮宇

3.2 Controlling the Learning

多数の自由パラメータと前頁に提示した学習アルゴリズム中には変更可能な部分がある．同様に ϵ の大きさについても前頁で紹介したが、これは急勾配をとる各ステップの大きさを決定するもので、p_{ij} と p'_{ij} が推定される時間の長さは学習過程に重要な影響を持つ.

p_{ij} と p'_{ij} を推定する実践的なシステムは推定の中で、いくつかのノイズを必要とするが、G のその値の中で「uphill steps」の原因を導く。

ネットワーク内の隠れ素子から G の中の局所最適解を作ることができ, これは (負債, 負担, 責務, 障害, 重荷,liability) を必要としない.

その推定の中でのノイズの影響は減少させることができ, 要望があれば, ϵ に小さな値を使ったり, より長い時間をかけて統計を集めたり, そして比較的簡単に G のその (最小値, 極小値,minimum) について焼きなまし探索を実装する.

目的関数 G は、どのようにして 2 つの確率分布が一致するか指定する測度である.

問題は次のような場合に起こる．もし, これまでに起きた可視素子 (を越える, 上の, について, に関して, 覆う,over) 可能なパターンの小さな部分集合だけを環境が指定したときだ.

(初期設定,default) で言及されていないパターンは発生する確率がゼロでなければならない。そして、0 度でない温度で実行されるボルツマンマシンが (ある程度の, いくつかの,certain) (設定,configurations) が起こり得ないと保障できる唯一の方法はそれらの (設定,configuration) に非常に高いエネルギーを与えることだが, それには (非常,infinitely) に大きな重みを必要とする.

非常に大きな重み (infinitely fine weight) について, この暗黙的な要求を避ける方法の 1 つは、ときおり, 雑音となる入力ベクトルを供給することだ.

(意訳) これは、「正しい」入力ベクトルに対して各ビットを反転させる小さな確率を持つ (過程, プロセス,proess) を経たフィルタリングによって供給できる。

それらのノイズの加わったベクトルは可視素子に固定される.

もし, ノイズが小さいのであれば, 正しいベクトルは統計を (支配する, 権威を振う, 見下ろす,dominate) だろう．しかし、全てのベクトルは発生するいくつかの機会をもっており、(それで, そこで, だから,then) 非常に大きなエネルギーは必要とならないだろう。

この「雑音固定,noisy clamping」テクニックは、ここで全ての例を (表現する, 提示する,present) ために使われた.

これは, とても上手くはたらくが、それには完全に満足せず、そして、

これまでに発生した可能性のある入力ベクトルのほんの一部 (とき, ならば, なので,when) が大きく成長しすぎたところから防ぐ重みの他の手法を (調査, 研究,investigate) した.

次節で示すシミュレーションでは、式 (10) によって暗示される (露骨な, 明かな,obvious) 最急降下法 (steepest descent method) の (変更, 改良,modification) を採用している.

Instread of changing w_{ij} by an amount proportional to $p_{ij} - p'_{ij}$, it is simply incremented by a fixed ‘weight-step‘ if $p_{ij} > p'_{ij}$ and decremented by the same amount if $p_{ij} < p'_{ij}$.

(に等しい, になる, 状態に達っする,amount)

w_{ij} の変化の代わりに $p_{ij} - p'_{ij}$ の比例によって,

これは単純な $p_{ij} > p'_{ij}$ の条件で固定された ‘weight-step‘の増加と $p_{ij} < p'_{ij}$ の条件で同じ量によって減少.

最急降下法 (steepest descent) と比べた、この手法の長所は G の 1 次、2 次の微分の中で、幅広い (変分, 変動, 変化,variations) に立ち向えることだ.

It can make significant progress on dimensions where G changes gently without taking very large divergent steps on dimensions where G falls rapidly and then rises rapidly again.

これは,G が急速に落ちる次元上でのとても大きな (発散ステップ,divergent step) を取ることなしで穏かに変化する G の次元上での重要な進捗と, そのとき再び急速に (上昇, 増加, 増大, 膨らむ,rise) よ

うにする.

そのような場合で、式 (10) 内の ϵ について適した数値は存在しない.

渓谷の穏かに傾斜した (谷底, 床,ravine) に沿った進捗を許す十分に大きなどんな値であっても

その渓谷の急勾配な面で発散による上下振動を引き起こすだろう.