統計的機械学習における量子アニーリング

佐藤 一誠[1], 田中 宗[2], 栗原 賢一[3], 宮下 精二 [4], 中川 裕志[5]

統計的機械学習は、過去に蓄積されたデータから何らかのルールを自動的に抽出する情報処理技術である。我々は、ルールの抽出過程、すなわち学習プロセスにおいて量子揺らぎを導入し効率的に学習する手法を開発した。我々のアルゴリズムは機械学習分野で幅広く使われている変分ベイズ法に基づいているため、変分ベイズ法が適用可能である従来研究に対して容易に量子揺らぎを導入できる。つまり、広範囲な応用分野に対して量子揺らぎを導入することを可能にした。

1 はじめに

統計的機械学習は、過去に蓄積されたデータを学習データとして未知の問題解決を機械的に行う情報処理技術である。学習データを未知の問題解決に利用するためには、データをどのような形で抽象化するかが重要である。例えば、我々人間は数学の問題を解く場合に、過去に解いた問題を公式または解法として抽象的に表現することで、新たな問題を解けるようになる。統計的機械学習では、確率モデルによってデータの性質を抽象化する。機械学習は、これまで多種多様な分野の理論を基に学習手法が提案されている。例えば、確率・統計、情報理論、理論計算機科学、そして物理学などが挙げられる。我々は量子情報理論に基づき、量子揺らぎの制御を学習のプロセスに応用するための理論構築を目指す。

データの特性を確率モデルによって抽象化する場合、「潜在変数」と呼ばれる確率変数が重要な役割を果たす。潜在変数は、データ間の隠れた類似性を数学的に取り扱うために導入される。例えば、我々人間が新たな問題を解く場合、過去の類似した問題を想起することで解決に至る場合が多々ある。しかし、どの問題と類似しているかは問題には書いていないため、問題間の類似性は非観測である。したがって、類似性の推定を行うことで問題解決を行う必要がある。統計的機械学習では、この類似性を問題と問題との間に潜む確率変数、すなわち潜在変数であると仮定し推定する。潜在変数の推定は、学習データに潜む性質を知ることにつながるため、データ解析手法としてもよく用いられる。

具体的な例として、ある文書集合をいくつかのカテゴリーに分類したいとする。特に、分類に関してなんら情報を与えずに自動的にいくつかのカテゴリーに分類するような問題を考える。文

[1]東京大学 情報基盤センター, E-mail: sato@r.dl.itc.u-tokyo.ac.jp
[2]東京大学大学院 理学系研究科, E-mail:shu-t@chem.s.u-tokyo.ac.jp
[3]Google, E-mail:kenichi.kurihara@gmail.com
[4]東京大学大学院 理学系研究科,E-mail:miya@spin.phys.s.u-tokyo.ac.jp
[5]東京大学 情報基盤センター, E-mail: n3@dl.itc.u-tokyo.ac.jp

書の分類では、「キーワード」が文字通り重要な役割を果たす。「キーワード」に基づくカテゴリー分類の難しさとして、どの単語が「キー」になるのかは与えられた文書集合に依存して変わることが挙げられる。したがって、どの単語がキーワードになるかは、データ集合から学習する必要がある。具体的には、「ある単語がキーワードとなるか (1)/ならないか (0)」という二値確率変数を潜在変数として導入し、文書集合から推定することで、各々の文書を自動的に分類することができる。新たな文書を処理する場合は、その文書のカテゴリーを推定することで、どのような内容かを類推することや、類似文書（推定された同一カテゴリの文書）の検索などを行うことができる。このようにデータに潜む性質を潜在変数の推定として浮き彫りにすることで統計的機械学習が可能となる。

2　確率的潜在変数モデルにおける量子アニーリング

機械学習は、多くの場合、最適化問題として定式化される。また、確率的潜在変数モデルの学習では、多数の局所解を持つ非線形最適化問題として定式化される。このような場合に、統計物理学的アプローチであるシミュレーテッド・アニーリング (SA : Simulated Annealing) (Kirkpatrick et al., 1983) が主に用いられる。SA は、温度を模したパラメータを導入し、熱揺らぎを制御する（温度を徐々に下げる）ことで、局所解を避けながら、より最適な解を探索する手法である。近年、局所解を含む最適化問題を解く手法として、量子揺らぎを用いた量子アニーリング (QA : Quantum annealing) が量子情報理論において注目を集めている (Kadowaki and Nishimori, 1998; Farhi et al., 2001; Santoro et al., 2002)。

我々は経路積分表式に基づく方法で QA を実装した。我々の用いた QA は、相互作用を持つ並列化 SA として考えることができる。これは、量子系を鈴木・トロッター展開 (Trotter, 1959; Suzuki, 1976) により古典系にマッピングすることで導出される。量子揺らぎは複数のプロセスにおける潜在変数間の相互作用として導入される。例えば、N 人の研究者を、いくつかのグループに分ける問題を考える。どの研究者がどのグループに属するのかを潜在変数として定義する。グループの分け方（状態）を σ で表現する (図 1 上図参照)。各グループ内での共著論文数が多いほど確率 $p(\sigma)$ が高くなるようなモデル化を行ったとする。目的は確率 $p(\sigma)$ を最大にする σ を求めることである。一般に機械学習では、初期状態を変えた m 個の SA プロセスを走らせ、最も確率の高い σ を解とする。具体的には、各 j プロセスで独立に以下の問題を解き、m 個の中で最も確率が高い状態を解とする。

$$\sigma_j^* = \operatorname*{argmax}_{\sigma_j} \log p(\sigma_j) \tag{1}$$

我々の提案する QA では、この m 個の SA 間で相互作用させながら探索を行う (図 1 下図参照)。ここで、 σ_j $(j = 1, \cdots, m)$ をそれぞれ j 番目のプロセスの状態とする。また、 $\sigma_{m+1} = \sigma_1$ となっている。f を相互作用関数とする。このような枠組みは、鈴木・トロッター展開を用いることで、数学的に導出された手法である。我々の提案する QA では、以下のような m 個のプロセスにおけ

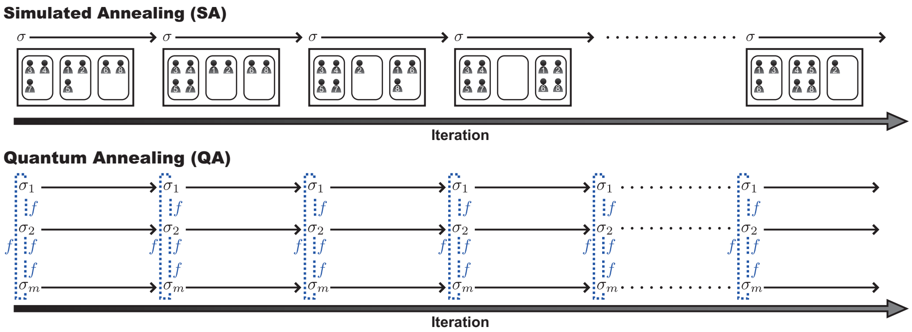

図 1: SA と QA：σ は潜在変数（この場合、8 人のグループ分け）を表している。

る状態 $\{\sigma_j\}_{j=1}^m$ に関する確率を最大にする問題を解く。

$$(\sigma_1^*, \sigma_2^*, \cdots, \sigma_m^*) = \underset{(\sigma_1, \sigma_2, \cdots, \sigma_m)}{\operatorname{argmax}} \log p_{\mathrm{QA}}(\sigma_1, \sigma_2, \cdots, \sigma_m) \tag{2}$$

ここで、$(\sigma_1, \sigma_2, \cdots, \sigma_m)$ は、潜在変数が m 個の値を同時に取った状態を表現し、この潜在変数の重ね合わせに対する確率分布を $p_{\mathrm{QA}}(\cdot)$ が定めている。図 1 の例では、σ_j $(j = 1, \cdots, m)$ は、各々異なるグループ分けの状態を示しており、QA は、m 個のグループ分け状態の重ね合わせ上の確率分布を基に解を探索していると考えられる。

最適化問題 (2) は、状態間の類似度を示す関数 $R(\sigma_1, \cdots, \sigma_m)$ を用いて、実際には以下のように書くことができる。

$$(\sigma_1^*, \sigma_2^*, \cdots, \sigma_m^*) = \underset{(\sigma_1, \sigma_2, \cdots, \sigma_m)}{\operatorname{argmax}} \sum_{j=1}^{m} \log p(\sigma_j) + f \cdot R(\sigma_1, \cdots, \sigma_m) \tag{3}$$

$f \cdot R(\sigma_1, \cdots, \sigma_m)$ が、潜在変数間の相互作用に相当する項である。またこの項は、実際には m 個の状態に関する制約を表現しているため、最適化問題 (3) は、最適化問題 (1) の制約付最適化問題としてみることができる。f が 0 の場合は、複数の最適化問題 (1) を独立に解くことに相当する。ここで、f 及び $R(\sigma_1, \cdots, \sigma_m)$ は、数学的に導出される。

我々は、最適化問題 (1) を近似的に解く手法として変分ベイズ法 (Attias, 1999) に着目し、最適化問題 (3) を近似的に解く量子アニーリング変分ベイズ法を提案した (Sato et al., 2009)。文書分類やトピック抽出に応用した結果、より効率的な学習ができることを確認した。変分ベイズ法は、自然言語処理、画像処理、音声処理、Web データ解析など多くの分野で用いられている汎用的な手法として知られている。したがって我々の手法は、変分ベイズ法で学習可能なモデルに対して適用可能であるため、これまで提案されてきた様々な応用分野で適用可能である。

3 おわりに

本稿では、潜在変数を含む確率モデルの学習において、量子揺らぎを導入する枠組みを紹介した。量子揺らぎは、統計力学的アプローチである熱揺らぎとは異なる揺らぎであるため、量子揺

らぎの制御は古典系の学習プロセスとは異なる学習プロセスとなる。このような量子系の学習プロセスが学習効率にどのような形で影響を与えるかについては未解明の問題であり、今後取り組んでいきたいと考えている。また、我々の手法は並列学習アルゴリズムであるため、大規模な並列数による学習効率の効果についても調べていく予定である。

謝辞

数値計算の一部は、東京大学物性研究所の共同利用スーパーコンピューターを利用いたしました。ここに感謝申し上げます。

参考文献

H. Attias. Inferring Parameters and Structure of Latent Variable Models by Variational Bayes. In K. B. Laskey and H. Prade, editors, *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 21–30, 1999.

E. Farhi, J. Goldstone, S. Gutmann, A. L. J. Lapan, and D. Preda. A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP -complete Problem. *Science*, 292:472–476, 2001.

T. Kadowaki and H. Nishimori. Quantum Annealing in the Transverse Ising Model. *Physical Review E*, 58:5355–5363, 1998.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.

G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car. Theory of Quantum Annealing of an Ising Spin Glass. *Science*, 295:2427–2430, 2002.

I. Sato, K. Kurihara, S. Tanaka, H. Nakagawa, and S. Miyashita. Quantum Annealing for Variational Bayes Inference. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, 2009.

M. Suzuki. Relationship between d-Dimensional Quantal Spin Systems and $(d+1)$-Dimensional Ising Systems – Equivalence , Critical Exponents and Systematic Approximants of the Partition Function and Spin Correlations –. *Progress of Theoretical Physics*, 56(5):1454–1469, 1976.

H. F. Trotter. On the Product of Semi-Groups of Operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959.