

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, 宮下精二², 中川裕志¹

東京大学 情報基盤センター¹, 東京大学 理学系研究科², Google³

我々人間があるテーマを「学習する」ことは、そのテーマに関する様々な情報を「未知の問題へ活用可能な形で抽象化する」ことであると言える。統計的機械学習は、蓄積された多くのデータや人の経験を数学的に抽象化することで、将来の予測を機械的に行うことを目指している。統計的機械学習では、確率モデルを用いてデータの性質をモデル化する。抽象化された情報は、確率モデルのパラメータとして表現される。機械学習では、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行っている。

データを統計的に抽象化する場合、統計的機械学習では、「潜在変数」と呼ばれる確率変数が重要な役割を果たす。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して解くことがある。このような類似性は、非観測である(どの問題と類似しているかは、問題には書いていない)。そのため、類似性の推定により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。文書の分類では、「キーワード」が文字通り重要な役割を果たす。しかし、どの単語が「キー」になるのかは問題によって変わることが、カテゴリーを分類することの困難さの一因として挙げられる。例えば「機械学習」という単語は、情報科学全般の文書集合を分類するキーワードになるが、機械学習に関する文書集合を分けるキーワードにはならない。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。この「キーワード」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。新たな文書进行处理する場合は、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。

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。機械学習における統計物理学的アプローチとしてよく知られているのはシミュレーテッド・アニーリング(SA: Simulated Annealing) (Kirkpatrick et al., 1983) である。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、量子情報理論では、別の揺らぎとして、量子揺らぎを用いた量子アニーリング(QA: Quantum annealing) が最適化問題を解く手法とし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。

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

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 ($j = 1, \dots, m$) 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。

$$(\sigma_1^*, \sigma_2^*, \dots, \sigma_m^*) = \operatorname{argmax}_{(\sigma_1, \sigma_2, \dots, \sigma_m)} \log p_{\text{QA}}(\sigma_1, \sigma_2, \dots, \sigma_m) \quad (2)$$

ここで、 $(\sigma_1, \sigma_2, \dots, \sigma_m)$ 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 $p_{\text{QA}}(\cdot)$ が定めている。図 1 の例では、 σ_j ($j = 1, \dots, m$) 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 $R(\sigma_1, \dots, \sigma_m)$ を用いて、実際には以下のように書くことができる。

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

$R(\sigma_1, \dots, \sigma_m)$ 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 $R(\sigma_1, \dots, \sigma_m)$ は、数学的に導出される。

我々は、確率的潜在変数モデルの学習手法として変分ベイズ法 (Attias, 1999) に着目し、量子揺らぎの効果を導入する手法を提案している (Sato et al., 2009)。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である。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確率的潜在変数モデルの学習に用いることができる。本発表では、文書分類、個人の音楽再生履歴の分析、Web ページ閲覧履歴の分析などに、提案手法を用いた場合の実験結果について紹介する。

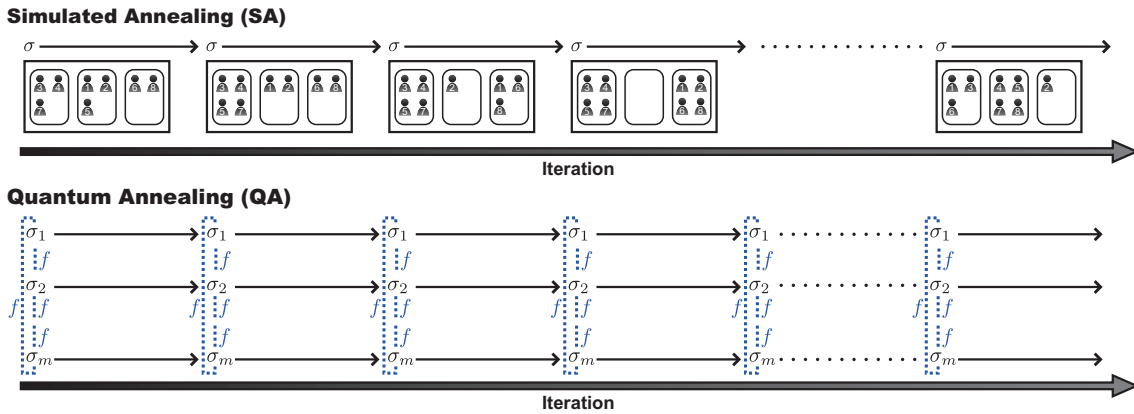


図 1: SA と QA : σ は潜在変数 (この場合、8 人のグループ分け) を表している。

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.