

量子アニーリングを用いた非負二値行列因子分解による レコメンデーション手法の検討

香月 諒大* 村上 公平* 矢実 貴志* 稲葉 陽子* 生田目 崇†

和文概要 近年、組合せ最適化問題を高速に解くためのアーキテクチャとして、量子アニーリングが注目されている。機械学習などの分野で量子アニーリングの活用方法について多数の研究がなされ始めているが、その多くがアルゴリズム等の手法の検討であり、課題解決的な実問題への検証が少ない。本稿では、実際のテレビの閲覧履歴データに対して、量子アニーリングを利用した非負二値行列因子分解を行い、視聴者の閲覧傾向の推定を行う手法の検討を行った。この推定手法はレコメンデーション等に対する応用が考えられ、その手法もあわせて提案を行う。数値実験を行った結果、推定を行うための学習データとなる閲覧情報が少ない状況下において、提案手法による推定性能が、既存手法による推定性能を上回っていることを確認した。

キーワード: 量子アニーリング, 組合せ最適化問題, 非負二値行列因子分解, 組合せ最適化, レコメンデーション

1. はじめに

近年、組合せ最適化問題を計算するためのアーキテクチャとして、量子アニーリングが注目されている。組合せ最適化問題は、実応用上の様々な数理モデルや機械学習の領域における基盤となっているが、問題の構造によってはNP-CompleteやNP-Hardと呼ばれる問題が存在する。これらの問題は多項式時間での計算が困難であり、計算量の観点から扱いにくいことが知られている。そのため、これらのクラスに属する大規模な問題に対しては、厳密な最適解の求解を一度諦め、様々なHeuristicsを加えるなどして計算の負荷を軽減し、できるだけ良質な解を導くといったアプローチがしばしば取られている。

量子アニーリングはそれらの手法と同様に、計算が難しいNP-Hardな組合せ問題に対して、Heuristicに解を導く手法として知られている[1]。現在、量子アニーリングを計算する専用マシンとして、D-Wave Systems Inc. のD-Wave 2000Q（以下、「D-Wave」とする）[2, 3]がある。D-Waveへの問題の実装には特徴があり、具体的には、解きたい対象の組合せ最適化問題と同様の構造を持つ物理系を構成し、その物理系に対するハミルトニアンが物理現象として最小化がなされた結果を取得して、組合せ最適化問題に対する解として利用する構成となっている。

量子アニーリングでは物理実験のハミルトニアンへのマッピングを行うので、求解対象となる組合せ最適化問題の構造にも限りがあり、イジングモデルが入力可能な形式となる。 $J \in \mathbb{R}^{N \times N}$, $\mathbf{h} \in \mathbb{R}^N$ に対して、イジングモデルの最小化問題は次のような式で表される。

$$\underset{\mathbf{x} \in \{-1, 1\}^N}{\text{minimize}} \quad \mathbf{x}^\top J \mathbf{x} + \mathbf{h}^\top \mathbf{x} + \text{const.} \quad (1.1)$$

活用先の拡大をテーマとして、機械学習等の基盤となる組合せ最適化問題の高速化を量子アニーリングによって図る研究も提案されている[4]。この先行研究では、D-Waveを用い

* 株式会社NTTデータ 東京都江東区豊洲3-3-9 豊洲センタービルアネックス

† 中央大学 理工学部経営システム工学科 東京都文京区春日1-13-27

た非負二値行列因子分解 (Non-negative/Binary Matrix Factorization, NBMF) のアルゴリズムを提案し、画像の基底となる特徴量の抽出タスクに対する性能の検証を行っている。 N 次元で表現されている M 枚の画像データ情報を保持した行列 $V \in \mathbb{R}^{N \times M}$ に対する NBMF は次のように表現される。

$$\underset{W \in \mathbb{R}_+^{N \times K}, H \in \{0,1\}^{K \times M}}{\text{minimize}} \quad \|V - WH\|_F^2 \quad (1.2)$$

NBMF を計算する際のアルゴリズムとして、交互最小二乗法 (Alternating Least Squares, ALS) [5] が知られている。ALS は決定変数 W, H を各ステップでいずれかを定数とみなして固定しつつ、交互に最小化を行う手法であるが、 H に対する部分問題は組合せ最適化問題となり、D-Wave を用いて計算を行っている。

一般に、 H に対して二値制約を置かない、非負値行列因子分解 (Non-negative Matrix Factorization, NMF) はより広く用いられており、NBMF と同様の画像の基底特徴量抽出 [6, 7] のみならず、次元圧縮やレコメンドの分野でも用いられている [8, 9]。NBMF はこの派生形であり、分解後の行列 H の要素が 0 と 1 しか取りえないため目的関数 (1.2) の最適値は劣化するが、少ない情報量でデータを表現することができるといったメリットがあると考えられる。行列 H に着目すると二値ラベリングをされたマルチラベル分類とみなすこともできる。また応用先として、画像の次元縮約による特徴抽出だけではなく、NMF と同様にレコメンドやクラスタリングでも活用できることが期待できる。

本研究では、量子アニーリングを利用した NBMF により、テレビ視聴履歴データを対象とした番組レコメンドーションの手法の提案と、その性能の検証を行う。まず、テレビ閲覧履歴に関する情報から、視聴者と番組に関する閲覧傾向を表す情報を抽出し行列形式で表現する。この行列に対して NBMF を適用することで、元の行列が持つ情報から重要なサンプルのみ抽出を行う。真の視聴者の好みを反映した閲覧傾向があると仮定しそれを推定するとき、ノイズを含んだテレビの閲覧データに対して、NBMF で情報の圧縮を行うことにより、視聴者の好みの閲覧傾向を推定できると考える。このような推定手法は W, H の双方を非負の連続値行列とする NMF においても用いられているが [8]、 H が疎性であることを仮定することで、組合せ最適化問題の求解速度だけではなく、推薦タスクとしての観点からみた推定性能も変更されることが期待される。

数値実験の結果、テレビ閲覧履歴の情報の多くが欠落している状況下において、提案手法は二乗平均平方根誤差 (Root Mean Squared Error, RMSE) の観点から NMF と比較したとき、推薦タスクとしての高い推定性能を持っていることを確認した。また、NBMF の計算速度の観点では、 H の解を求めるために生じる部分問題の求解において、既存手法である SA と比較したところ、計算速度の向上を確認することができた。

2. 非負二値行列因子分解 (NBMF) について

本章では、先行研究 [4] で提案されている、D-Wave による NBMF のアルゴリズムを紹介する。NBMF の目的関数は (1.2) で与えられている。 W, H はいずれも決定変数であるが、式 (1.2) に対し ALS で計算を行う上では、 W, H のうち片方のみを決定変数とみなし、もう片方を定数と見做した部分問題を交互に解く。

まず、 $W \in \mathbb{R}_+^{N \times K}$ を計算する際には射影勾配法 [5] を用いて、

$$W \leftarrow W \circ (VH^T \circ WHH^T) \quad (2.1)$$

という更新を施す。 \odot, \ominus はそれぞれアダマール積, 商を表す。一方, H に対する更新は, 次のような組合せ最適化問題の解で与えられる。

$$\underset{H \in \{0,1\}^{K \times M}}{\text{minimize}} \|V - WH\|_F^2 = \sum_{i=1}^M (H_i^\top W^\top W H_i - 2V_i^\top W H_i + V_i^\top V_i) \quad (2.2)$$

更に決定変数の固定を行って, $i = 1, \dots, M$ を各々別々に解くことにより, $H_i \in \{0,1\}^K$ に対する最適化問題とする。この決定変数に対してアフィン変換を施すことによって, $\{-1,1\}^K$ が探索を行う対象の空間となり, イジングモデルに帰着することができる。このイジングモデルの最小化問題に対して, D-Wave を用いて計算を行う。

3. 提案手法

3.1. テレビ閲覧履歴行列の構成

レコメンデーションの対象とするテレビの視聴者数を N , 番組の数を M とする。テレビ閲覧履歴行列 $V \in \mathbb{R}^{N \times M}$ には, その (i, j) 成分に対して, 視聴者 i が番組 j を閲覧した時間(分)を実績として保存する。

3.2. クラスタリング

NBMF は, 視聴者のテレビ閲覧実績や傾向データに基くが, 実際にレコメンデーションを行う場合, 視聴者や番組に付随する有益な情報が多々ある。例えば, 視聴者に対して性別や年齢, 職業による偏りがあり, テレビ番組に対しては, 放映されているテレビ局・時間帯などにより大別することができる。そこで, 予めこれらの静的な情報をもとにしてユーザや番組を k-means 法によりクラスタリングしておき, それらのクラスに対してテレビ閲覧履歴行列を構成する。

3.3. レコメンデーション

テレビ閲覧履歴行列は, 過去の閲覧情報の実績を元に作成される行列であるが, 番組数が非常に多く, 視聴者は全ての番組情報を知らないまま番組閲覧している状況などから, テレビ閲覧履歴行列の値と, 真の視聴者の好みを反映した閲覧傾向との間に乖離が生じると考えられる。

テレビ閲覧履歴行列 V_{record} に対して, 2章で記述した NBMF を用いて得られた行列を $W_{\text{estimate}}, H_{\text{estimate}}$ とする。推定閲覧傾向行列を $V_{\text{estimate}} = W_{\text{estimate}} H_{\text{estimate}}$ として定めたいので, 差分行列 D を次のように定める。

$$D = V_{\text{estimate}} - V_{\text{record}} \quad (3.1)$$

行列 D の (i, j) 要素 $D_{i,j}$ について, $D_{i,j} \geq 0$ の場合には, 実際には閲覧傾向が高いにもかかわらず, 閲覧実績が少ないとみなして, 視聴者 i に対して, 番組 j をレコメンドを行う。逆に, $D_{i,j} < 0$ の場合には, 過剰評価の状態であると判定を行う。

4. 数値実験

4.1. 対象とするデータ

本論文では, 経営科学系研究部会連合協議会主催・平成 30 年度データ解析コンペティションで提供された株式会社ビデオサーチの VR-CUBIC データを用いる。本データには, 2016 年から 2017 年において, 視聴者が閲覧した番組に関する分単位のデータ, 及び視聴者

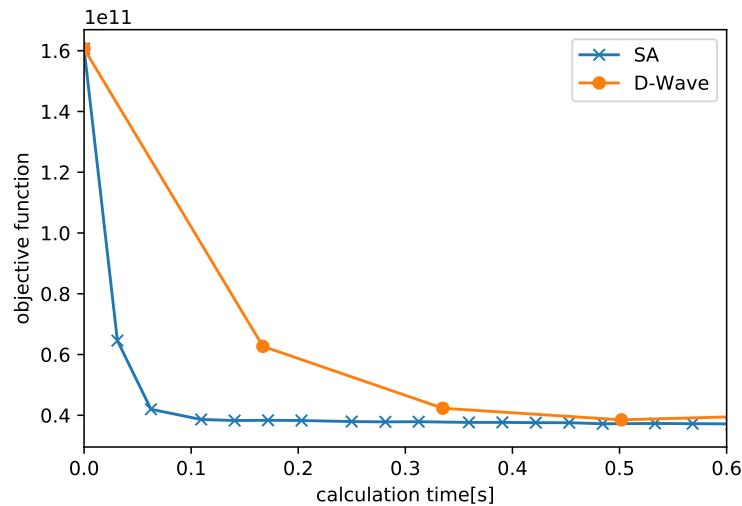


図 1: V_{record} に対し, $K = 8$ における目的関数 (1.2) の時間推移

と番組のプロフィールデータが含まれている. ザッピングのノイズを防ぐ目的で閲覧時間の閾値処理を行った後, 視聴者の閲覧履歴と番組情報を統合し, 要素に視聴時間を持つテレビ閲覧履歴行列を作成し, 数値実験の入力 V_{record} とする.

4.2. 数値実験諸元

クラスタリングのサイズは, 視聴者数が $N = 10$, 番組の数が $M = 15$ となるよう設定し, ALS の反復時間は 10 回とした. D-Wave のバージョンは DW_2000Q_VFYC_5 を指定しており, (2.2) の場合, 64bit が求解サイズの上限となる. なお, 以下の数値実験結果を測定する際には, 通信時間は除外の対象とし, 秒数を測定した. 投入する問題の係数幅は $[-1, 1]$ 区間となるよう定数倍を行ったうえで, chain_strength=1.2 とし, その他のパラメータはデフォルトの設定を用いた [10]. また, 比較対象とする SA は同 D-Wave Ocean SDK 内の neal ライブラリのデフォルトの設定を利用している [11].

4.3. NBMF の推定性能の評価

性能評価にあたり, V_{record} に対して, 恣意的に σ ($0 < \sigma < 1$) の割合のデータにマスキングを施した行列 V_{masked} を作成する [9]. 具体的には, V_{record} の全要素のうち, ランダムに確率 σ でマスキング対象とする. マスキング対象の行列の要素には, 同一行のマスキングされていない対象だけからなる行列要素の値の平均を取り平滑化を施す. (σ が大きく, ある行のすべての要素がマスキング対象となった場合には, 更に列方向へ平滑化を施す). 本研究では NBMF の分布推定性能指標として, $V_{\text{record}}, V_{\text{masked}}$ 間の RMSE である $\|V_{\text{record}} - V_{\text{masked}}\|_2$ を利用し評価を行った. また結果の平均化のために, NMF, NBMF どちらについてもマスキング対象のランダム性に対して 10 回 RMSE を評価した平均値を計算し記載している.

4.4. 数値実験結果

4.4.1. 計算時間の評価

NBMF における部分問題 (2.2) の計算手法を SA 及び D-Wave と変更した場合における, 目的関数 (1.2) の最小化の速度を比較する. V_{record} に対して, $\sigma = 1, K = 8$ としたときの計算速度は図 1 のようになった.

各反復において, SA, D-Wave 共に類似の目的関数値を辿るが, SA はより短時間で反復

が進んでいる。ALS の 1 回の反復あたり，両手法共に部分問題 (2.2) が 15 回固定で呼び出されるため，SA と D-Wave の計算時間の差は部分問題 (2.2) の求解時間にそのまま起因する。そこで，表 1 では，部分問題 (2.2) の部分問題のサイズを変更し，D-Wave と SA の計算時間の特性を比較した。計算時間測定のためのデータとして， $N = 50$ ， $M = 50$ ， $[0, 1]$ 区間内で各値をランダムに与えた $V \in \mathbb{R}^{N \times M}$ を準備し， K を変更しながら評価している。初期値における目的関数 (1.2) は 11189.36 であり，表 1 には ALS1 反復目に得られた目的関数 (1.2) も併せて記載する。計算時間は， $M = 50$ としたために生成される 50 個の部分問題 (2.2) の平均とした。

表 1: K に対する ALS の 1 反復目における部分問題 (2.2) の計算時間及び目的関数 (1.2)

	アルゴリズム	$K = 10$	$K = 20$	$K = 30$	$K = 40$
計算時間 (秒)	SA	0.00201	0.00376	0.01458	0.01602
	D-Wave	0.01003	0.01011	0.01014	0.01017
目的関数 (1.2)	SA	202.001	171.150	146.102	118.489
	D-Wave	223.813	222.377	256.819	231.729

K のサイズが大きくなるにつれて，SA の計算時間は徐々に拡大している一方で，D-Wave の計算時間はほぼ一律であり， $K = 40$ において，D-Wave は短時間で 1 反復を実現している。目的関数 (1.2) の観点では，SA に比べて劣化しているように見受けられる。

4.4.2. 推定性能の評価

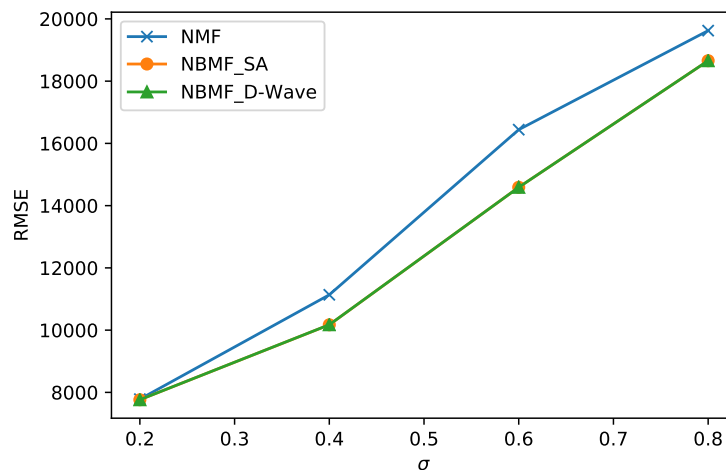


図 2: σ の変更に対する $K = 8$ での NMF と NBMF の RMSE

図 2 には σ を変更した場合の各手法の RMSE を示しており，D-Wave と SA の間の RMSE の差はほぼ認められない。これは両手法ともに NBMF の目的関数 (1.2) としており，また部分問題の呼び出し元となる ALS アルゴリズムが共通であるため，収束先の局所解が類似しているためだと考えられる。一方，元の行列に対してマスキングされる割合に着目したときには，特に σ が 0.4 から 0.6 の近傍である場合において，RMSE の観点から NBMF は NMF に比べて推定性能が優れている。

5. 結論

本稿では、実際のテレビ閲覧のデータに対して、D-Wave を利用した NBMF により閲覧傾向の推定を行い、レコメンデーションを行う手法の提案を行った。

数値実験の結果、NBMF は NMF と比較して、マスキング前のテレビ閲覧履歴行列の値の推定性能が RMSE の観点で優れていることを示した。また、計算時間の観点では、 K のサイズが大きくなるにつれて、組合せ最適化問題の既存手法である SA は計算時間が拡大するのに対して、D-Wave の場合には固定の短時間での計算が可能である。

本研究の課題点として、D-Wave はハードウェアとして実装されているビット数が不足や、クラウドで公開されている関係上、通信遅延のために検証可能な問題規模に限界があることが挙げられる。また、計算時間の観点では D-Wave は最適性が必ずしも保証されないアルゴリズムであるため、問題規模が大きく拡大した際にもスケールするか検討が必要である。得られた行列 H のスパースな構造をうまく活用して、対応する番組をマルチラベル分析するなど、得られた結果の活用方法についても検討の余地があると考えられる。

参考文献

- [1] Kadowaki, T., and Nishimori, T.: Quantum annealing in the transverse Ising model, *Physical Review E*, Vol. 58, No. 5, pp. 5355–5363, 1998.
- [2] Johnson M. W., Amin, M. H. S., Gildert, S., Lanting, T., Hamze, F., Dickson, N., et al.: Quantum annealing with manufactured spins, *Nature*, Vol. 473, pp. 194–198, 2011.
- [3] Gibney, E.: D-Wave upgrade: How scientists are using the world’s most controversial quantum computer, *Nature*, Vol. 541, pp. 447–448, 2017.
- [4] O’Malley, D., Vesselinov, V. V., Alexandrov, B. S., and Alexandrov, L. B.: Nonnegative/Binary matrix factorization with a D-Wave quantum annealer, *PLoS ONE*, Vol. 13, pp. 1–12, 2018.
- [5] Lin, C. J.: Projected gradient methods for nonnegative matrix factorization, *Neural Computation*, Vol. 19, No. 10, pp. 2756–2779, 2007.
- [6] Eee, D. D. L., and Seung, H. S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol. 401, pp. 788–791, 1999.
- [7] Eee, D. D. L., and Seung, H. S.: Algorithms for non-negative matrix factorization, *Proceedings of the 13th International Conference on Neural Information Processing Systems (NIPS’00)*, pp. 535–541, 2000.
- [8] Koren, Y., Bell, R. M., and Volinsky, C.: Matrix factrization techniques for recommender systems, *Computer*, Vol. 42, pp. 30–37, 2009.
- [9] Bennett, J., and Lanning, S.: The Netflix prize, *Proceedings of KDD Cup and Workshop 2007*, pp. 3–6, 2007.
- [10] D-Wave Solver Properties and Parameters Reference, https://docs.dwavesys.com/docs/latest/_downloads/09-1169A-G_D-Wave_Solver_Properties_Parameters_Reference.pdf (2019年11月13日現在)
- [11] dwave-neal, <https://docs.ocean.dwavesys.com/projects/neal/en/latest/> (2019年11月13日現在)