Harmonic mean similarity based quantum annealing for k-means

Takano, Jo
Omori, Toshiaki

(Citation)
Procedia Computer Science,144:298-305

(Issue Date)
2018

(Resource Type)
journal article

(Version)
Version of Record

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 144 (2018) 298–305

Procedia Computer Science

www.elsevier.com/locate/procedia

INNS Conference on Big Data and Deep Learning 2018

Harmonic mean similarity based quantum annealing for *k*-means

Jo Takano, Toshiaki Omori*

Department of Electrical and Electronic Engineering, Graduate School of Engineering, Kobe University, Kobe, Hyogo, 657-8501, Japan

Abstract

Clustering is an important machine learning approach in analyzing big data. In this study, we propose a clustering method using quantum annealing (QA) based on harmonic average of purity and inverse purity. By using harmonic average of purity and inverse purity, we introduce the effect of quantum noise for clustering algorithm while considering both quality of clusters and quality of categories. Using benchmark data, we show the effectiveness of the proposed QA clustering based on the harmonic average of purity and inverse purity.

Keywords: Machine learning algorithm, Multi-dimensional data analysis, Clustering, Quantum annealing

1. Introduction

Clustering is an important machine learning approach in analyzing big data [6]. There are many situations where there is no label information. Clustering makes it possible to find latent structure and patterns from the data without label information. Since data generated from the same class have similar features, multiple data from the same class tend to have some pattern and those data is distributed locally in the feature space. The *k*-means method, which is simple and is based on the distance in feature space, has been used extensively so far. However, the *k*-means clustering has a disadvantage, which comes from large initial value dependency; in complicated cases where the size and the number of clusters are large, it is difficult to obtain an optimal solution by falling into a local solution.

In order to overcome this problem of *k*-means, deterministic approaches using initial value optimization and probabilistic approaches have been proposed. *k*-means++ [2] is a deterministic one which considers initial value optimization to better values. On the other hand, a probabilistic approach based on simulated annealing (SA) [3] has been proposed in order to avoid local optimum solution using thermal noise.

In this study, we propose a clustering method using quantum annealing (QA) based on harmonic average of purity and inverse purity [1]. Here we use the effect of quantum noise for clustering algorithm [9] while considering quality

* Corresponding author.
E-mail address: omori@eedept.kobe-u.ac.jp

1877-0509 © 2018 The Authors. Published by Elsevier Ltd.

10.1016/j.procs.2018.10.531

of clusters (purity) and quality of categories (inverse purity). Quantum annealing [7] is a method for minimizing the objective function by using quantum noise and is known to deal with combinatorial optimization problems such as traveling salesman problem [10]. Here we consider quantum annealing based on Suzuki-Trotter expansion, which leads to multiple simulated annealing with an interaction between systems. Using benchmark data, we evaluate the effectiveness of the proposed QA clustering based on the harmonic average of purity and inverse purity by comparing existing SA clustering and QA clustering [9, 13] with the proposed QA clustering.

2. SA-*k*-means

In this section, we describe a concept of SA and then explain k-means based on SA. The SA is a probabilistic optimization method for estimating the state that minimizes the objective function E. In the SA, we consider state change based on probability ratio from the current state to the candidate state. When the objective function decreases $\{\Delta E = E_{next} - E_{present} < 0\}$, the state change is always accepted. When the objective function increases, we accept the state change with the acceptance probability $\exp[-\Delta E/T]$ according to the increment ΔE and the temperature T. A wide area search tends to be performed with high temperature T, and a local search tends to be performed with low temperature. In the SA, we start optimization with high temperature T and we decrease the tempreerature to simulate thermal annealing. However SA is known to fall into a local minimum [12].

Here we explain the method using simulated annealing for the k-means method. We assign certain data x_j to the cluster c_i stochastically, where $X = \{x_1, x_2, ..., x_n\}$ is a set of data and $c = \{c_1, c_2, ..., c_k\}$ is a set of cluster center vectors. In this probabilistic framework, we consider the probability depends on the distance r_{x_j,c_i} between data x_j and cluster center vector c_i and the temperature T. The algorithm of SA k-means is described below.

1. Allocate cluster randomly to all data $X = \{x_1, x_2, ...x_n\}$
2. Update all cluster center vectors c_i
3. The cluster is assigned to all data $X = \{x_1, x_2, ..., x_n\}$. Here using the probability ratio $\exp[-\beta r_{x_j,c_i}]$. r_{x_j,c_i} is the Euclidean distance between x_j and c_i
4. Update β
5. Repeat steps 2.3.4

Here $\beta = 1/T$ is the inverse temperature, which controls SA. Namely, the temperature T is a gradually decreased from initial high temperature. Global search is performed at high temperature while local search is performed at low temperature. Thus, by adapting SA to k-means, we aim to avoid the problem of locally optimal solution due to initial value dependence.

3. QA-*k*-means

In this section, we propose clustering method based on quantum annealing using harmonic mean of two different similarities. First, we explain a basic framework of clustering using QA in subsection 3.1. Then we describe our proposed formulation of QA k-means based on harmonic mean of purity and inverse purity in subsection 3.2. We formulate the similarity to be used for quantum annealing.

3.1. Loss function

Fig. 1 shows a concept of quantum annealing. Here, we define an expression of state using binary vector in order to define loss function used in the framework of quantum annealing. Let σ_i be a one-hot-vector of size K that represents which cluster the i th data is allocated to. When a cluster center vector set is expressed using a matrix $c = [c_1, c_2, ..., c_K]$, $c\sigma_i^{(j)}$ express a cluster center vector to which i-th data belongs. The cluster center vector and data are allocated so that the distance between cluster center vector $c\sigma_i^{(j)}$ and the i-th data x_i becomes small.

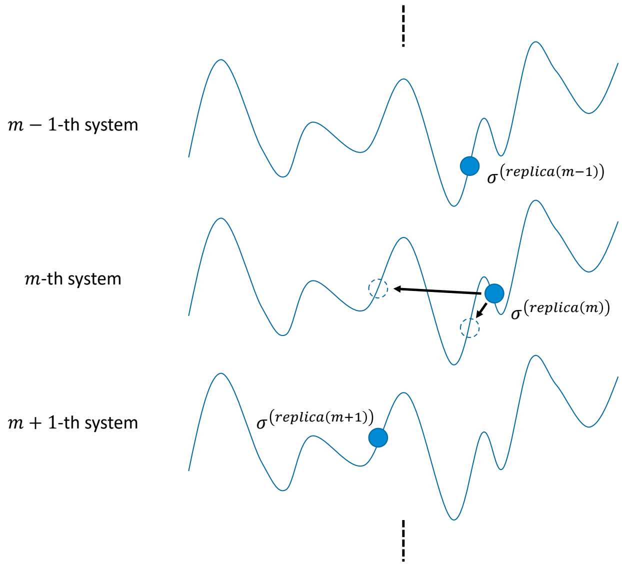

Fig. 1. The concept of quantum annealing. By considering multiple systems with interaction, we realize an effective search using quantum effects that can pass through barriers. In this figure, interactions to m-th system is explicitly shown by arrows in for simplicity. $\sigma^{(replica(m))}$ interacts to adopt state which is similar to $\sigma^{(replica(m-1))}$ and $\sigma^{(replica(m+1))}$. Detailed explanations are given in subsection 3.2.

Thus, the loss function (sum of squared error, SSE) can be formulated as follows:

$$\text{Loss}(\sigma^{(j)}) = \sum_{i=1}^{N} \left(x_i - c\sigma_i^{(j)}\right)^2 \tag{1}$$

where we assign the cluster to all the data and let the combination be $\sigma^{(j)} = \otimes_{i=1}^{N} \sigma_i^j$. $\otimes$ means Kronecker delta. $\sigma^{(j)}$ is a one-hot-vector whose size is K^N, where K^N is the total number of possible cluster assignments. We estimate cluster allocation $\sigma^{(j)}$ and cluster vector c that minimize this loss function.

3.2. Proposed QA k-means based on harmonic mean

To realize clustering using quantum annealing, we consider that Hamiltonian consists of classical term H_c and quantum term H_q as follows:

$$H = H_c + H_q \tag{2}$$

The first term H_c is a matrix for the loss of the problem and is a diagonal matrix as follows:

$$H_c = \text{diag}\left(\text{Loss}\left(\sigma^{(1)}\right), \text{Loss}\left(\sigma^{(2)}\right), \cdots, \text{Loss}\left(\sigma^{(K^N)}\right)\right) \tag{3}$$

The second term H_q is a matrix for quantum fluctuation as follows:

$$H_q = -\Gamma \sum_{i=1}^{N} \sigma_i^x \tag{4}$$

where σ_x is Pauli's matrix as follows:

$$\sigma_i^x = \left(\otimes_{x=1}^{i-1} \mathbb{E}_K\right) \otimes (\mathbb{1}_K - \mathbb{E}_K) \otimes \left(\otimes_{x=i+1}^{N} \mathbb{E}_K\right) \tag{5}$$

where $\mathbb{E}_K$ is a $k \times k$ identity matrix, $\mathbb{1}_K$ is a $k \times k$ matrix where every element is equal to one. Here, Γ is transverse magnetic field, corresponding to a parameter for quantum annealing.

Using the Hamiltonian shown in Eq. (2), the probability of appearance of the state $\sigma^{(j)}$ is as follows [11]:

$$\begin{aligned} p(\sigma^{(j)}) &= \frac{{\sigma^{(j)}}^{\mathrm{T}} \exp\left[-\beta H\right] \sigma^{(j)}}{\sum_{\sigma^{(k)}} {\sigma^{(k)}}^{\mathrm{T}} \exp\left[-\beta H\right] \sigma^{(k)}} \\ &\propto {\sigma^{(j)}}^{\mathrm{T}} \exp\left[-\beta H\right] \sigma^{(j)} \equiv \tilde{p}(\sigma^{(j)}) \end{aligned} \tag{6}$$

In stochastic cluster assignment, we use probability ratio of $\tilde{p}(\sigma^{(j)})$.

Here we consider approximate using Suzuki-Trotter formula

$$\exp\left[A_1 + A_2\right] \simeq \lim_{M\to\infty} \left[\exp \frac{A_1}{M} \exp \frac{A_2}{M}\right]^M \tag{7}$$

This formula enables us to decompose exponential function of sum of matrices into the product of exponential functions of individual matrix. By assuming that M is sufficiently large, we can separate classical and quantum Hamiltonians as follows:

$$\begin{aligned} \tilde{p}(\sigma^{(j)}) &= {\sigma^{(j)}}^{\mathrm{T}} \exp\left[-\beta(H_c + H_q)\right] \sigma^{(j)} \\ &\simeq {\sigma^{(j)}}^{\mathrm{T}} \exp\left[-\frac{\beta}{M} H_c\right]^M \exp\left[-\frac{\beta}{M} H_q\right]^M \sigma^{(j)} \end{aligned} \tag{8}$$

After this transformation, we have M classical systems instead of quantum systems. The superposition of quantum states is replaced by the search in a plurality of classical systems with interactions between classical systems.

By performing analytical calculation, we can obtain following probability for the state of $\sigma^{(j)}$ [5, 4].

$$\tilde{p}\left(\sigma^{(replica(m))} = \sigma^{(j)}\right) \propto \exp\left[-\frac{\beta}{M} \mathrm{Loss}(\sigma^{(j)}) + f(\beta, \Gamma) \sum_{n=1}^{N} \left(\sigma_n^{(replica(m-1))} \sigma_n^{(j)} + \sigma_n^{(j)} \sigma_n^{(replica(m+1))}\right)\right] \tag{9}$$

where $f(\beta, \Gamma)$ is a function of β, Γ. As shown in Fig. 1, this can be thought of as multiple simultaneous simulated annealing with interactions among the systems. This represents the probability that the state of the m-th system is $\sigma^{(j)}$

where total number of systems in M. The state of the m-th system seems likely to be similar to the state in the system of $(m-1)$-th system and $(m+1)$-th system.

In Eq. (9), β and Γ are annealing parameters to control in QA. As shown in factor $\sum_n^N \sigma_n^{(replica(m-1))}\sigma_n^{(j)}$, we calculate the number of matched clusters assigned by comparing m and $(m-1)$-th systems. Here we denote similarity between systems by $s(\sigma^{(a)}, \sigma^{(b)})$. This is generalized and further expressed by the following expression using the function $\beta(t)$, $f_{interaction}(t)$ representing the schedule of annealing.

$$\tilde{p}(\sigma^{(replica(m))} = \sigma^{(j)}) \propto \exp\left[-\beta(t) r_{x_i,c_k} + f_{interaction}(t)\left\{s\left(\sigma^{(replica(m-1))}, \sigma^{(k)}\right) + s\left(\sigma^{(k)}, \sigma^{(replica(m+1))}\right)\right\}\right] \tag{10}$$

Here we consider three kinds of evaluation indices for clustering [1]. The first evaluation index, purity, is defined as follows:

$$\text{Purity} = \sum_i \frac{|C_i|}{N} \max_j \text{Precision}(C_i, L_j) \tag{11}$$

where C_i is the cluster assignment of the system and, L_j is correct cluster assignment. Precision is expressed using C_i and L_j as follows:

$$\text{Precision}(C_i, L_j) = \frac{|C_i \cap L_j|}{|C_i|} \tag{12}$$

Naturally, purity is the degree to which clusters are moistened with the same L_j and represents the uniqueness in the cluster. However, this index does not reward the assignment of correct labels to the same cluster. In the extreme case, assigning one cluster per one data maximizes purity.

To overcome this problem in purity, we introduce a second evaluation index, inverse purity, an index of similarity with properties different from purity.

$$\text{Inverse Purity} = \sum_i \frac{|L_i|}{N} \max_j \text{Precision}(C_i, L_j) \tag{13}$$

Namely, inverse purity is the degree to which the same L_j is concentrated in one cluster and represents the convergence to a certain cluster. In this study, we propose a clustering using the harmonic averaged indicator of these two evaluation indices, F-measure, as third evaluation index,

$$\text{F} = \left(\alpha \frac{1}{\text{Purity}} + (1-\alpha)\frac{1}{\text{Inverse Purity}}\right)^{-1} \tag{14}$$

where α is a constant and is set to be $0 < \alpha < 1$. In this study, we employ one of three indices as the Similarity s in Eq. (10) in order to realize clustering using QA.

4. Evaluating the Proposed Method

In this section, we show the effectiveness of clustering using QA using the proposed method f-measure and compare it with other annealing methods.

4.1. Setting

We use a clustering benchmark dataset called A3 in the experiment [8]. The data has 50 clusters and high frustration as shown in Fig. 2. The control parameters used in Eq. (10) are set as follows. The inverse temperature β is controlled the convergence guarantee schedule $\beta(t) = \beta_0 \log t$. The transverse magnetic field Γ is controlled the schedule of $\Gamma(t) = \Gamma_0 \frac{M}{\beta}\frac{\mathcal{T}}{t}$ [12]. where $\mathcal{T}$ is the total number of iterations.

4.2. Result

In this subsection, we perform clustering using proposed and existing methods. Fig. 3 shows the results of clustering using such as k-means++, SA, purity-based QA, and f-measure-based QA. For each clustering method, the SSE loss function is plotted. As shown in Fig. 3, the k-means and k-means++ methods which are deterministic algorithms, show fast convergence, but it falls into a local solution. In the QA method, the SSE becomes lower than the k-means++ and the SA method around 270 iterations. In particular, f-measure-based QA shows lower SSE than purity QA near convergence. Furthermore, f-measure-based QA can be executed with almost the same time as purity-based QA, since f-measure-based QA only increases the calculation of inverse purity in Eq. (10).

5. Conclusion

In this study, we proposed a clustering method using quantum annealing (QA) based on harmonic average of purity and inverse purity. Here we used the effect of quantum noise for clustering algorithm. Using benchmark data, we shown the effectiveness of the proposed QA clustering based on the harmonic average of purity and inverse purity.

Introduction of quantum annealing based on harmonic mean similarity may have advantages in other clustering methods such as Gaussian mixture model. We leave this as a future work.

Acknowledgements

This work is partially supported by Grants-in-Aid for Scientific Research for Innovative Areas Initiative for High-Dimensional Data driven Science through Deepening of Sparse Modeling [JSPS KAKENHI Grant No. JP25120010] and for Scientific Research [JSPS KAKENHI Grant No. JP16K00330], and a Fund for the Promotion of Joint International Research (Fostering Joint International Research [JSPS KAKENHI Grant No. JP15KK0010] from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] Amigó, E., Gonzalo, J., Artiles, J., Verdejo, F., 2009. A comparison of extrinsic clustering evaluation metrics based on formal constraints. Information retrieval 12, 461–486.
[2] Arthur, D., Vassilvitskii, S., 2007. k-means++: The advantages of careful seeding, in: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics. pp. 1027–1035.
[3] Černỳ, V., 1985. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. Journal of optimization theory and applications 45, 41–51.
[4] Chandra, A.K., Das, A., Chakrabarti, B.K., 2010. Quantum quenching, annealing and computation. volume 802. Springer Science & Business Media.
[5] Das, A., Chakrabarti, B.K., 2005. Quantum annealing and related optimization methods. volume 679. Springer Science & Business Media.
[6] Everingham, M., Zisserman, A., Williams, C.K., Van Gool, L., Allan, M., Bishop, C.M., Chapelle, O., Dalal, N., Deselaers, T., Dorkó, G., et al., 2006. The 2005 pascal visual object classes challenge, in: Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment. Springer, pp. 117–176.

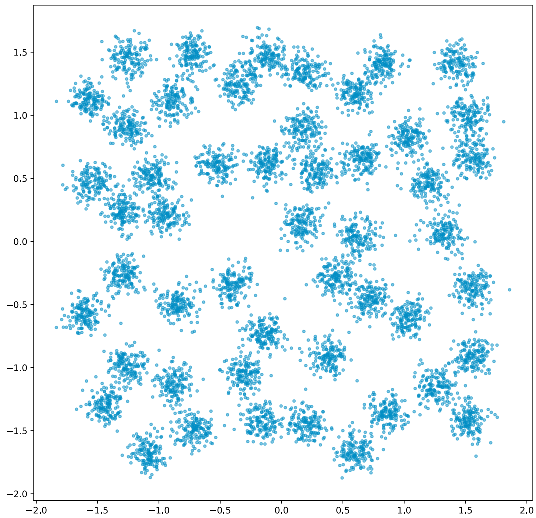

Fig. 2. Benchmark dataset for clustering.

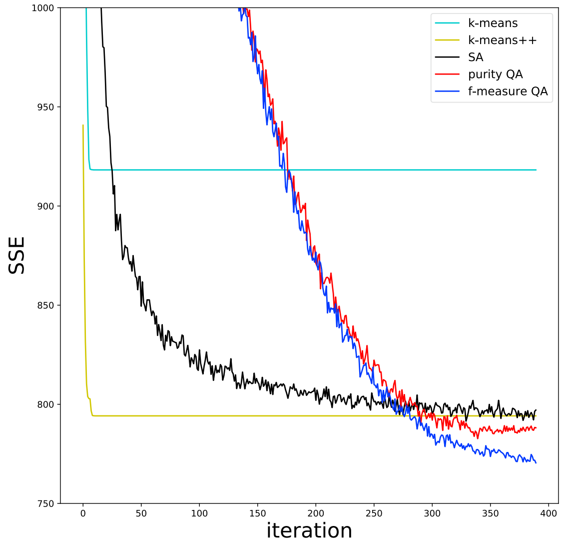

Fig. 3. Sum of squared errors (SSE) curves obtained by *k*-means, *k*-means++, SA, purity QA and f-measure QA.

[7] Kadowaki, T., Nishimori, H., 1998. Quantum annealing in the transverse ising model. Physical Review E 58, 5355.
[8] Kärkkäinen, I., Fränti, P., 2002. Dynamic local search algorithm for the clustering problem. University of Joensuu.
[9] Kurihara, K., Tanaka, S., Miyashita, S., 2009. Quantum annealing for clustering, in: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, AUAI Press. pp. 321–328.
[10] Martoňák, R., Santoro, G.E., Tosatti, E., 2004. Quantum annealing of the traveling-salesman problem. Physical Review E 70, 057701.
[11] Nishimori, H., 2001. Statistical physics of spin glasses and information processing: an introduction. volume 111. Clarendon Press.

[12] Sato, I., Tanaka, S., Kurihara, K., Miyashita, S., Nakagawa, H., 2013. Quantum annealing for dirichlet process mixture models with applications to network clustering. Neurocomputing 121, 523–531.

[13] Sun, L.X., Xu, F., Liang, Y.Z., Xie, Y.L., Yu, R.Q., 1994. Cluster analysis by the k-means algorithm and simulated annealing. Chemometrics and intelligent laboratory systems 25, 51–60.