Ising-based Combinatorial Clustering using the Kernel Method

○Masahito Kumagai\textsuperscript{1}, Kazuhiko Komatsu\textsuperscript{2}, Masayuki Sato\textsuperscript{1}, Hiroaki Kobayashi\textsuperscript{1}

\textsuperscript{1} Graduate School of Information Sciences, Tohoku University

\textsuperscript{2} Cyberscience Center, Tohoku University
Backgrounds

• Ising machines
  • Solve combinatorial optimization problems
  • E.g., D-wave Quantum Annealer, Fujitsu Digital Annealer

• Combinatorial clustering
  • Minimize the sum of the intra-cluster degree of similarities among data points
    \[ W(C) = \frac{1}{2} \sum_{a=1}^{K} \sum_{C(i) = C_a} \sum_{C(i') = C_a} d(x_i, x_{i'}) \]
  • Euclidean distance as degree of similarity
    • Difficult to perform complex distribution data
Binary clustering using the kernel method

- Ising-based binary clustering using the Kernel method [Bauckhage, 2018]

\[
H = \frac{1}{2} \sum_{i,j=0}^{N-1} Q_{i,j} s_i s_j . \quad s_i = \begin{cases} 
+1 & (C(i) = C_0) \\
-1 & (C(i) = C_1)
\end{cases}
\]

- Kernel method
  - Map the data onto a high-dimensional feature space
  - Transformed data can be used for Euclidean distance-based clustering
  - Gram matrix: \( Q_{i,j} \), Kernel function: \( k(x_i, x_j) \)

\[
Q_{i,j} = k(x_i, x_j) - \frac{1}{n} \sum_{l} k(x_i, x_l) - \frac{1}{n} \sum_{k} k(x_k, x_j) + \frac{1}{n^2} \sum_{k,l} k(x_k, x_l) . \quad k(x_i, x_j) = \exp\left( - \frac{1}{2\sigma^2} d(x_i, x_j) \right)
\]

Kernel clustering based on the Ising model has been proposed only when the number of clusters is 2.
Objective and Approach

• Objective
  • To achieve combinatorial clustering using the kernel method based on the Ising model for two or more clusters

• Approach
  • Propose a general QUBO (Quadratic Unconstrained Binary Optimization) formulation of combinatorial clustering based on the kernel method
  • Represent clustering results using the one-hot encoding
An Overview of the Proposed Method

• Decision variables using One-Hot Encoding

\[ q_a^i = \begin{cases} 
1 & (C(i) = C_a) \\
0 & (C(i) \neq C_a) 
\end{cases} \]

<table>
<thead>
<tr>
<th>(a)</th>
<th>(i)</th>
<th>(0)</th>
<th>(1)</th>
<th>(\ldots)</th>
<th>(N-2)</th>
<th>(N-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>\ldots</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>\ldots</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>(K-2)</td>
<td>0</td>
<td>0</td>
<td>\ldots</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(K-1)</td>
<td>0</td>
<td>0</td>
<td>\ldots</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Objective function

• Objective function
  • The sum of intra cluster elements of the Gram matrix

\[ H_{\text{objective}} = \frac{1}{2} \sum_{i,j=0}^{N-1} Q_{i,j} \sum_{a=0}^{K-1} q_a^i q_a^j \]

• Eliminate redundancy because of the symmetric Gram matrix

\[ H_{\text{objective}} = \frac{1}{2} \sum_{i,j=0}^{N-1} Q_{i,j} \sum_{i<j}^{N-1} q_i^i q_j^j + \sum_{i<j}^{N-1} Q_{i,j} \sum_{a=0}^{K-1} q_a^i q_a^j \]

minimum value when all binary variables are 0
→ all the data do not belong to any cluster
One-Hot constraint

- One-Hot constraint
  - All the data belong to only one cluster
    \[ \exists! q_a^i \in \{q_a^0, \ldots, q_a^{N-1}\} \quad \text{s.t. } q_a^i = 1 \]

- Two methods to satisfy the One-Hot constraint
  - Minimize the combined function of both objective and constraint function
    \[
    H = \frac{1}{2} \sum_{i,j=0}^{N-1} Q_{i,j} \sum_{a=0}^{K-1} q_a^i q_a^i + \sum_{i<j}^{N-1} \sum_{a=0}^{K-1} q_a^i q_a^j + \lambda \sum_{i=0}^{N-1} \left( \sum_{a=0}^{K-1} q_a^i - 1 \right)^2
    \]
    - Objective
    - Constraint
  - \( \lambda \) is given by the method of the Lagrange multiplier
    - Sufficiently large \( \lambda \) is needed for violation of the One-Hot constraint
    - Since the influence of the objective function is small, the clustering quality decreases

- Minimize only the objective function while externally defined One-Hot constraint
  - Obtain high quality results [Kumagai, 2021]
Experimental Environments

• Clustering methods
  • Proposed: Combinatorial clustering using the kernel method
  • Conventional: Euclidean distance-based combinatorial clustering

• Hardware
  • SX-Aurora TSUBASA
    • Intel Xeon 6126 (CPU)
    • NEC Vector Engine Type 20B (VE)
  • Aurora Simulated Annealing (ASA)
    • Externally defined the One-Hot constraint and searches for solutions that allow multiple bit flips
    • sweeps: 100, inverse temperature: 0.1 to 50, the number of parallel temperature: 8

• Data sets
  • Generated using the artificial data generation functions of scikit-learn v0.23.2

• Metric to evaluate clustering quality
  • Adjusted Rand Index (ARI)
Clustering results

Both conventional and proposed method can appropriately clustering data with equal variances

Only proposed method successfully clustering anisotropic and complex data
This is because data is transformed into linearly separable high-dimensional data by the kernel method
Qualities for different datasets

- **Blobs**
  - Conventional > Proposed
  - Proposed is also high at 0.91
  - Linearly partitioned without any transformation

- **Aniso, Moons, and Circles**
  - Proposed >> Conventional
  - Data transformation by the kernel method is effective for complex data
Qualities for different parameters

To obtain high quality clustering results, it is necessary to select appropriate parameter $\sigma$ of the kernel function $k(x_i, x_j) = \exp(-\frac{1}{2\sigma^2}d(x_i, x_j))$.

Selecting the appropriate parameter improves the quality of the proposed method.
Clustering results for many clusters

- The proposed method can handle two or more clusters
- The proposed method has better clustering performance for complex data than the conventional method even when the number of clusters is large
Qualities for different clusters

- Changing the number of circles and the number of clusters

- Proposed shows higher ARI than Conventional

- ARI of Proposed decreases as the number of clusters increases
  - This is because local optimal solutions are obtained by Simulated Annealing
  - Necessary to appropriately adjust the parameters for SA
Conclusions and Future Work

• Backgrounds
  • Combinatorial clustering using the kernel method can perform clustering of the complex data
  • However, kernel clustering based on the Ising model has been proposed only when $K = 2$

• Proposal
  • A general QUBO formulation of combinatorial clustering based on the kernel method by using the One-Hot encoding

• Results
  • The proposed method can handle two or more clusters
  • The qualities of the clustering results is significantly higher than that of the conventional method on the complex data
  • The proposed method can obtain high-quality results by selecting the appropriate parameter

• Future Work
  • Comparison with other state-of-the-art methods
  • Combining other Ising-based clustering with Ising-based kernel clustering
Thank you for listening

If you have any questions, please contact me
masahito.kumagai.t5@dc.tohoku.ac.jp