Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.98549517

LABEL CORRESPONDENCE OPTIMIZATION MODEL

Ms Urvashi Soni[1], Dr. Sunita Dwivedi[2]
[1] Research Scholar MCRPV Bhopal, [2] Associate Professor, MCRPV Bhopal

ABSTRACT

This research explores the clustering ensemble problem, which tries to aggregate various base clustering to generate performance that is superior to that of the individual one. As a weighted linear combination of the connective matrices from various base clustering, the existing clustering ensemble methods typically create a co-association matrix, which indicates the pairwise similarity between samples. The resulting Co-association matrix is then adopted as the input of a pre-existing clustering algorithm, such as Meta clustering. The co-association matrix, meanwhile, could be dominated by weak base clustering, leading to weak performance. In order to address the issue, we suggest a new matrix of similar label approximation based approach in this study. We specifically create a Cohesive Matrix, which comprises a small but highly reliable set of links between samples, by examining whether two samples are grouped to the same cluster with various base clustering. The Cohesive Matrix and the Co-association matrix are then stacked to create a three-dimensional Matrix, whose label correspondence quality is further investigated to convey the Cohesive Matrix's information to the Co-association matrix and create a more accurate co-association matrix. We frame and effectively solve the proposed approach as a smooth confined optimal solution. Comparing the proposed model to 11 state-of-the-art approaches, experimental results over 7 benchmark data sets demonstrate that it delivers a breakthrough in clustering performance. To the best of our knowledge, this study to investigate the potential of a clustering ensemble using a matrix of similar label, which is fundamentally different from other methods. Last but not least, our technique just has one easily adjustable parameter.

Introduction:

The purpose of the vital but difficult unsupervised task of clustering is to divide a set of samples into cohesive subgroups. Many applications, including recommender systems, community identification, and picture segmentation, can be framed as a clustering problem. Numerous clustering methods have been proposed throughout the years, such as K-means and Meta clustering. Hierarchical clustering, matrix factorization, Gaussian mixture models, and others. Since each technique has benefits and disadvantages of its own, no technique can always dominate another. A clustering approach often includes a few hyper-parameters, on which its effectiveness greatly depends. Additionally, it can be challenging to adjust the hyper-parameters, and some algorithms, like K-means, are quite sensitive to initialization. These dilemmas make it more challenging to select the best clustering technique for a normal clustering task. In order to do this, the clustering ensemble method was developed. Its goal is to construct a consensus clustering that performs better than the base clustering given a set of base clustering created by various methods or the same method with various hyper-parameters and initializations. Clustering ensemble is more challenging than supervised methods, as voting and other popular supervised methods procedures cannot be easily applied to clustering

ensemble whenever sample labels are not present. Existing techniques often learn a pairwise association matrix from the base clustering before applying pre-made techniques like Meta clustering to the resulting matrix to obtain the final clustering result. We basically introduced by U.Soni in 2022 classify the existing approaches into two groups based on the way the pairwise co-association matrix is generated:

1. In order to develop a pairwise similarity matrix, the first class of approaches treats the base clustering as new multilayer perceptron (as illustrated in Fig. 1-A). For instance, clustering ensemble was defined as a convex matrix of similar label representation issue by Gao in 2016, searched for a dense affinity matrix for clustering ensemble using a Frobenius norm self-representation model.

2. The Co-association matrix (shown in Fig. 1-C), which summarizes the co-occurrence of samples in the same cluster of the base clustering, is the foundation of the second category of approaches. Co-association matrices were first introduced by Fred and Jain, and ever since then, they have gained popularity as a key fundamental technique in clustering ensembles. Hypothetically, weighted K-means clustering reduces computing complexity by theoretically bridging the co-association based technique to weighted K-means clustering.

Numerous sophisticated approaches for building co-association matrices have recently been developed. For instance, suggested a locally weighted Co-association matrix after taking the uncertainty of each base clustering. Cluster-wise similarities were employed to improve the conventional co-association matrix. A self-paced learning method for the Co-association matrix was suggested. View the specifics in the next section. We note that the former co-association matrices of earlier research has variations in weighted linear combination of the connective matrices (as shown in Fig. 1-B) from various base clusters. When some base clustering performs poorly, they dominate the Co-association matrix and significantly lower clustering performance. In order to improve the Co-association matrix from a global standpoint, we provide a unique restricted matrices of similar label approximation (MSA) model in this study. We then build a Cohesive Matrix, whose element determines whether or not two samples are from the same cluster in all of the base clustering, as illustrated in Fig. 1-D. We next combine the standard Co-association matrix with the Cohesive Matrix to create the 3-dimensional (3-D) Matrix depicted in Figure 1—E, which is made up of additional matrices with approximated labels. The new proposal can refine the Co-association matrix by examining the label correspondence, propagating the Cohesive Matrix's highly reliable information to the co-association matrix. This refined Co-association matrix is then used as the input of available clustering method to produce the final clustering result. Practically speaking, the new proposal is an alternate iterative solution to a convex optimization problem. We test the suggested model against seven benchmark data sets and 11 clustering ensemble techniques. The range of research prove that the suggested model performs noticeably better than existing approaches. We believe this study to investigate the impact of similar label matrices on clustering ensembles.

Related Work:

Notation: We use letter $\boldsymbol{\mathcal{A}}$ to indicate array-matrix (*an algebraic object that describes a multi-linear relationship between groups of algebraic objects related to a vector space*.), **A** to denote matrices, **a** and a to denote vectors and scaler. Let $\boldsymbol{\mathcal{A}}$ (i, j, k) stand for the 3-D Matrix (i, j, k) th element. **a**(i) indicates the i-th entry of vector **a**, while **A**(i, j) indicates the(i, j)-th member of matrix **A**. Matrix $\boldsymbol{\mathcal{A}}$ A's i-th frontal slice is denoted as $\boldsymbol{\mathcal{A}}$. (:, :, i).

Matrix tier: In this study we use a matrix nuclear norm computed by label single value decomposition (kilmer 2013) to measure the label. Especially the 3-d Matrix $\boldsymbol{\mathcal{A}} \in R^{n_1 * n_2 * n_3}$ can stated as

$$\boldsymbol{A} = \boldsymbol{U} * \boldsymbol{S} * \boldsymbol{V}^{\boldsymbol{T}} \qquad (1)$$

Here$\boldsymbol{U} \in \boldsymbol{R}^{n_1 * n_2 * n_3}$ and $\boldsymbol{V} \in \boldsymbol{R}^{n_1 * n_2 * n_3}$are the two analogous label, $\boldsymbol{S} \in \boldsymbol{R}^{n_1 * n_2 * n_3}$ is an diagonal label, $*$ and ^{T} denote label product and label transpose, respectively

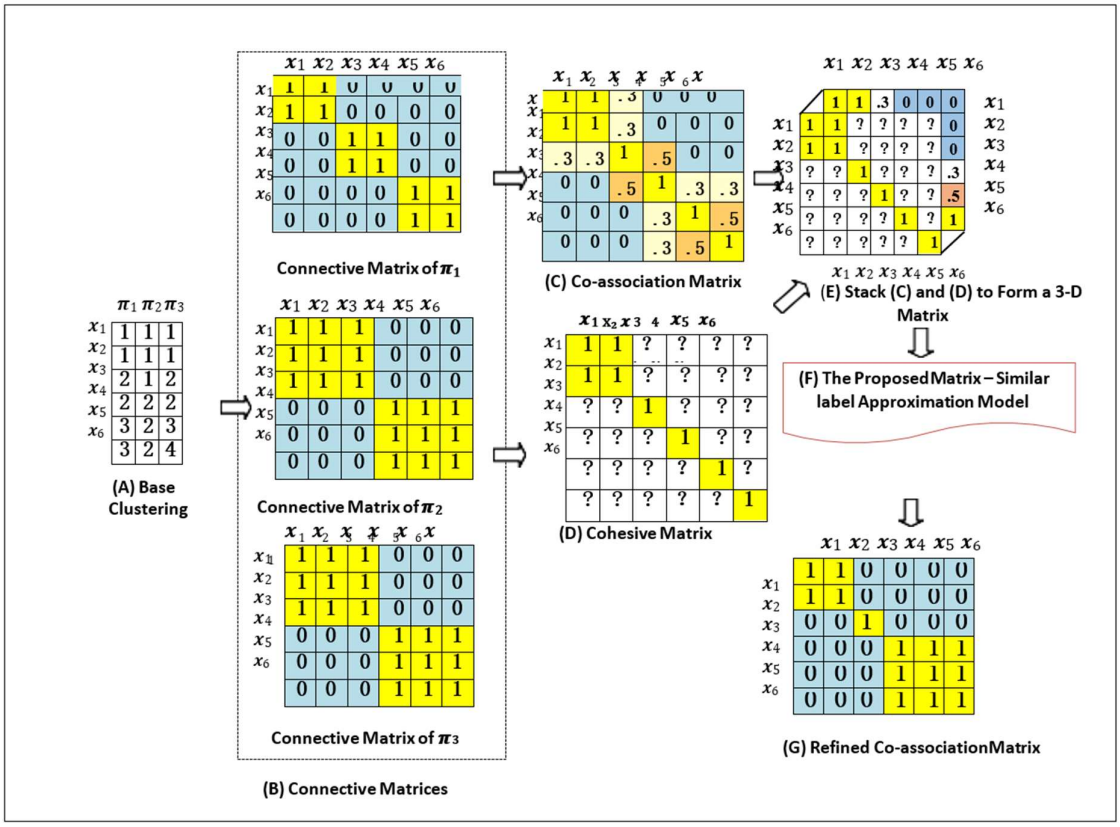

Figure- 1: Using three base clusters, represented by π_1, π_2, and π_3, and six input samples, represented by x_1, x_2, x_3, x_4, x_5, x_6, as an example, the proposed approach is demonstrated. The limited but very reliable information present in the Cohesive Matrix can be used to improve the co-association matrix's quality by examining the label correspondence of the created 3-D matrix.

Algorithm 1 l-SVD of a 3-D matrix (Zhang 2014)

Input: 3-D Matrix $\boldsymbol{\mathcal{A}} \in R^{n_1 * n_2 * n_3}$

1: Perform FFT on $\boldsymbol{\mathcal{A}}$, i.e., $\boldsymbol{\mathcal{A}}_{f}$ = fft($\boldsymbol{\mathcal{A}}$, [ ], 3);

2: for k = 1: n3; do

3: Perform SVD on each frontal slice of Af , i.e.,

[**U,S,V**]=SVD($\boldsymbol{\mathcal{A}}_f$ (:, :, k)) ;

4: $\mathbf{U}_f$ (:, :, k) = **U**, $\mathbf{S}_f$ (:, :, k) = **s**, $\mathbf{V}_f$ (:, :, k) = **V**;

5: end

6: Perform inverse FFT on $\mathbf{U}_f$, $\mathbf{S}_f$ and $\mathbf{V}_f$, i.e., **U** =

ifft($\mathbf{U}_f$, [], 3), **S** = ifft($\mathbf{S}_f$, [], 3) and **V** = ifft($\mathbf{V}_f$, [], 3);

Output: **U**, **S** and **V**.

The complete definitions of the matrix-related operators described above may be found in (Zhang 2014). The L-SVD form of a matrix can be derived with the fast Fourier transform (FFT is nothing but computation of discrete Fourier transform in an algorithmic format, where the computational part will be reduced.) efficiently as shown in Algorithm 1 because the matrix product can be computed in the Fourier domain efficiently (Kilmer 2013). The matrix nuclear norm is described by Zhang (2014) as the sum of the absolute values of the diagonal entries of $\boldsymbol{S}$, i.e.

$$\|\boldsymbol{\mathcal{A}}\|_{\circledast} = \sum_{i=1}^{min(n_1,n_2)} \sum_{k=1}^{n_3} |\boldsymbol{S}(i,j,k)|. \qquad (2)$$

$\circledast$ $\boldsymbol{is}$ APL Functional Symbol Circle Star

Formulation of Clustering Ensemble: Given a data collection $\mathcal{X} = [x_1, x_2, \dots x_n] \in \mathbb{R}^{d*n}$ containing n samples, each sample $x_i \in \mathbb{R}^{d*1}$, and m base clustering $\prod = [\pi_1, \pi_2, \dots \pi_m] \epsilon \mathbb{R}^{n*m}$ where each base clustering $\pi_i \in \mathbb{R}^{n*1}$ is an n-dimensional vector with the j-th element $\pi_i(j)$ representing the cluster The cluster indicators in various base clustering for clustering ensemble are often unique. A fictitious example with 6 samples and 3 base clustering is shown in Fig. 1-A. The goal of a clustering ensemble is to combine several m base clustering in order to outperform each one separately.

Proposed Method:

The preceding techniques, which create a Co-association matrix as the linear combination of connective matrices, are susceptible to some subpar base clustering. To do this, we suggest a brand-new approach based on matrix of similar labels approximation that may be used to improve the original co-association matrix.

Problem Formulation:

First, we build a Cohesive Matrix (as seen in Fig. 1-D), which checks if two samples are clustered to the same category under every base clustering, in order to refine the co-association matrix. It is important to note that the Cohesive Matrix's components are extremely trustworthy data that we might extrapolate from the base clustering. In particular, using the Co-association matrix, we might immediately obtain the Cohesive Matrix $M \epsilon \mathbb{R}^{n*n}$, i.e.

$$M(i,j) = \begin{cases} \mathbf{1}\ \boldsymbol{if}\ \boldsymbol{A(i,j)} = \mathbf{1} \\ \mathbf{0}\ \boldsymbol{otherwise} \end{cases} \quad (\mathbf{3})$$

The Cohesive Matrix and the Co-association matrix are then stacked to create the 3-D matrix $\boldsymbol{\mathcal{P}} \epsilon \mathbb{R}^{n*n*2}$, where $\boldsymbol{\mathcal{P}}(:,:,1) = \boldsymbol{M}$ and $\boldsymbol{\mathcal{P}}(:,:2) = \boldsymbol{A}$. The produced matrix should ideally be minimal because the members of both the Cohesive Matrix and the Co-association matrix indicate the pairwise similarity between samples. We attempt to complement the zero elements with reference to the non-zero ones and the Co-association matrix because the non-one elements of $\boldsymbol{M}$ are constrained but express the very reliable similarity between samples. The elements of the co-association matrix, on the other hand, are dense but have numerous wrong connections, and we attempt to improve it by deleting the faulty connections, which are represented by $\boldsymbol{E} \epsilon\ \mathbb{R}^{n*n}$, by using the data from the Cohesive Matrix. Additionally, each frontal slice of $\boldsymbol{\mathcal{P}}$ should be symmetric, and the elements of $\boldsymbol{\mathcal{P}}$ should be constrained to the range $[0, 1]$. The proposed method is formally defined as a label correspondence optimization problem, written as

$$min_{\boldsymbol{\mathcal{P}},\boldsymbol{E}} \|\boldsymbol{\mathcal{P}}\|_{\circledast} + \lambda\ \|\mathbf{E}\|_F^2 \quad (4)$$

Where $\boldsymbol{\mathcal{P}}(:,:,\mathbf{2}) = \mathbf{0}$ is placed on $\boldsymbol{E}$ to prevent trivial solutions and $\lambda = 0$ is the coefficient to balance the error matrix. By improving Eq. (4), it is anticipated that the Cohesive Matrix will be supplemented according to the information from the Co-association matrix at the same time as the restricted but highly reliable information in $\boldsymbol{M}$ is propagated to the co-association matrix.

We can get a refined Co-association matrix $\boldsymbol{\mathcal{P}}(:,:,\mathbf{2})$ after solving the issue in Eq. (4), with the optimal solution. The final clustering result can then be produced by using any pairwise similarity-based clustering methods on $\boldsymbol{\mathcal{P}}(:,:,\mathbf{2})$. In this article, we examine two widely used clustering techniques: agglomerative hierarchical clustering (Ng, Jordan, and Weiss 2002) and Meta clustering (Fred and Jain 2005).

Numerical Solution:

Based on the imprecise Augmented Lagrangian method, we suggest an optimization technique to solve Eq (4). (Jia, Kwong, and Hou 2018). To work with the limited and symmetric restrictions on $\boldsymbol{\mathcal{P}}(:,:,\mathbf{1})$and $\boldsymbol{\mathcal{P}}(:,:,\mathbf{2})$respectively, we first define two auxiliary matrices $\boldsymbol{B}, \boldsymbol{C}\ \epsilon \mathbb{R}^{\ n*n}$, and Eq. (4) can be equally expressed as

$$\boldsymbol{argmin}_{\boldsymbol{\mathcal{P}},\boldsymbol{E},\boldsymbol{B},\boldsymbol{C}} \|\boldsymbol{\mathcal{P}}\|_{\circledast} + \lambda\ \|\mathbf{E}\|_{\boldsymbol{F}}^{\mathbf{2}} \quad (\mathbf{5})$$

The augmented Lagrangian form we apply three Lagrange multipliers $\wedge_1$,$\wedge_2$ $and\ \wedge_{\mathbf{3}}\ \boldsymbol{\epsilon}\ \mathbb{R}^{\boldsymbol{n*n}}$

Where the penalty coefficient is $\mu > 0$. Then, it can be made more efficient by solving the following four subproblems alternately and iteratively, where only one variable is changed while the others are left fixed.

The $\boldsymbol{\mathcal{P}}$ sub-problem:

$\boldsymbol{\mathcal{P}}$ is expressed as after the unwanted terms have been removed. And expressed as

$$\boldsymbol{argmin}_{\boldsymbol{\mathcal{P}}} \frac{\mathbf{1}}{\mu} \|\boldsymbol{\mathcal{P}}\|_{\circledast} + \frac{\mathbf{1}}{\mathbf{2}} \|\boldsymbol{\mathcal{P}} - \boldsymbol{\mathcal{T}}\|_{\boldsymbol{F}}^{\mathbf{2}}, \quad (\mathbf{6})$$

Using the soft-threshold operator of the matrix singular values, it has a closed form solution, according to +Zhang (2014). Algorithm 1's computation of **FFT** and **SVD** on the frontal slices of the input 3-D matrix $\boldsymbol{\mathcal{T}}(:,:,\boldsymbol{i})$ and its FFT version $\boldsymbol{\mathcal{T}_f}$ $(:,:.\boldsymbol{i})$, respectively,

emphasizes the frontal slices' label correspondence. On the other hand, we hope to benefit from the adjustment between the initial Co-association matrix and the Cohesive Matrix. To obtain the L-SVD representation, we therefore execute FFT and SVD on the lateral slices of the matrix $\boldsymbol{\mathcal{T}}(:,\boldsymbol{i},:)$ and $\boldsymbol{\mathcal{T}_f}(:,\boldsymbol{i},:)$ respectively.

The $\boldsymbol{E}$ subproblem:

Without the irrelevant terms, the $\boldsymbol{E}$ subproblem becomes:

$$\boldsymbol{min_E} \succ \|\boldsymbol{E}\|_F^2 + \frac{\mu}{2}\left\|\boldsymbol{\mathcal{P}}(:,:,\mathbf{2}) + \boldsymbol{E} - \boldsymbol{A} + \frac{\Lambda_2}{\mu}\right\|_{\boldsymbol{F}}^{\mathbf{2}} \quad \textbf{(7)}$$

Putting the derivative of above Eq to 0 will result the global minimum because it is a complex quantity of $\boldsymbol{E}$.

The $\boldsymbol{B}$ sub-problem:

The B sub-problem is written as

$$\boldsymbol{min_B}\ \frac{\mu}{2}\left\|\boldsymbol{B} - (\boldsymbol{\mathcal{P}}(:,:,\mathbf{1}) + \frac{\Lambda_1}{\mu})\right\|_{\boldsymbol{F}}^{\mathbf{2}} \qquad \textbf{(8)}$$

According to Jia (2020d), has an ideal element-wise solution and is a symmetric and bounded restricted least.

The $\boldsymbol{C}$ sub-problem:

Without a set of element-wise equality constraints, the $\boldsymbol{C}$ sub-problem is the same as the B sub-problem and optimal solution of it. Update $\boldsymbol{\Lambda_1, \Lambda_2, \Lambda_3\ and}\ \mu,\ \boldsymbol{\mu_{max}}$ .The Lagrange multipliers and μ are updated. $\boldsymbol{\mu_{max}}$ Is upper bound and μ is initially set to 0.0001 (Liu 2019). The halting conditions for Algorithm 2 total numerical solution are $\mathbf{max}(\|\mathbf{B} - \boldsymbol{\mathcal{P}}(:,:,\mathbf{1})\|_\infty, \|\mathbf{C} - \boldsymbol{\mathcal{P}}(:,:,\mathbf{2})\|_\infty, \|\mathbf{A} - \mathbf{E} - \boldsymbol{\mathcal{P}}(:,:,\mathbf{2})\|_\infty) < \mathbf{10^{-8}}$ with $\|.\|_\infty$ is the highest possible value of a matrix's absolute values.

Experiment:

We performed numerous tests to assess the suggested hypothesis.

Algorithm 2 Numerical solution to Eq. (4)

Base clustering matrix π is entered

$\boldsymbol{\mathcal{P}}, \boldsymbol{E}, \boldsymbol{B}, \mathbf{C}$ and $\mu_{\max}$ are initialized to 0 and 10^8 respectively

1. Create the Co-association matrix $\boldsymbol{A}$;
2. To construct the Cohesive Matrix $\boldsymbol{M}$,;
3. While not converged do
4. update $\boldsymbol{\mathcal{P}}$;
5. update $\boldsymbol{E}$;
6. update $\boldsymbol{B}$;
7. update $\boldsymbol{C}$;
8. $\boldsymbol{\Lambda_1, \Lambda_2, \Lambda_3 and}\ \mu$;

9. To check the convergence requirements.

10. Finish while producing the revised Co-association matrix $\boldsymbol{\mathcal{P}}(:,:,\mathbf{2})$ as output**.**

Data Sets:

We selected 7 frequently used data sets, including BinAlpha, Multiple features (MF), MNIST, Semeion, CalTech, Texture, and ISOLET, in accordance with current clustering ensemble works (Huang, Wang, and Lai 2018; Huang, Lai, and Wang 2016; Zhou, Zheng, and Pan 2019). In accordance with Huang, Wang, and Lai (2018), we chose 5000 samples at random from MNIST and used the subset in the experiments. We also used 20 typical categories out of 101 categories for CalTech, which we referred to as CalTech20.

Generation of Base Clustering:

In accordance with Huang, Wang, and Lai (2018), we first created a pool of 100 potential base clustering for each of the data sets by using the K-means technique, where $\boldsymbol{K}$ is a random number with a value that varies between $[2, \sqrt{n}]$, where n is the total number of input data samples.

Methods under Comparison:

We contrasted the proposed model with 11 state-of-the-art clustering ensemble techniques, such as the TA-CL, TA-SL, and PTGP (Huang, Lai, and Wang 2016), LWSC, and LWGP (Huang, Wang, and Lai 2018), E-HC and E-MC (Huang 2018), DREC (Zhou, Zheng, and Pan 2019), SPCE (Zhou 2020), and sMLCA (2018), SEC (2017 Liu.et.al). The authors offer the codes for each of the approaches that were compared. The proposed model is designated as sMLCA and OMC, respectively, and is provided with agglomerative hierarchical clustering and Meta clustering to produce the final clustering result.

Evaluation Metrics:

We used the clustering accuracy (ACC), normalized mutual information (NMI), purity, adjust rand index (ARI), F1-score, precision, and recall measures to assess clustering performance. The values of all the measures are up-bounded by 1, and for all of them, a higher number indicates greater clustering performance. You may find the comprehensive definitions of those measures in (Zhang 2020; Jia 2020b).

Experiment Settings:

We choose 10 base clustering at random for each data set from the candidate base clustering pool, then we applied several clustering ensemble techniques on the chosen base clustering. We performed the random selection 20 times in order to reduce the impact of the chosen base clustering, and we then reported the average performance over those 20 iterations. We set the hyper-parameters for the comparison algorithms in accordance with the original papers. Unless no suggested values were available, we thoroughly examined the hyper-parameters and used those that produced the best results. The single hyper-parameter in the suggested model, which was set to 0.002 for all the data sets.

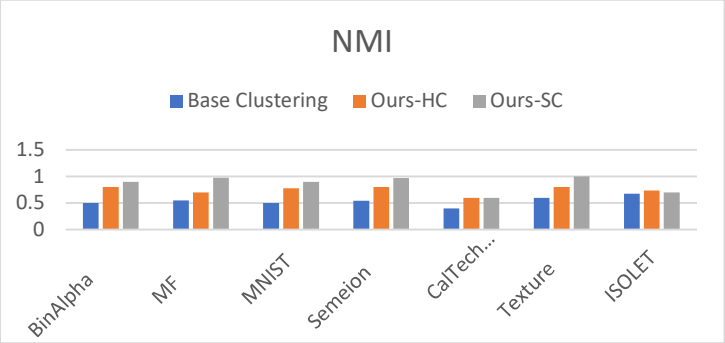

Figure 2: The NMI of our methods against the average NMI of the base clustering in the candidate base clustering pool.

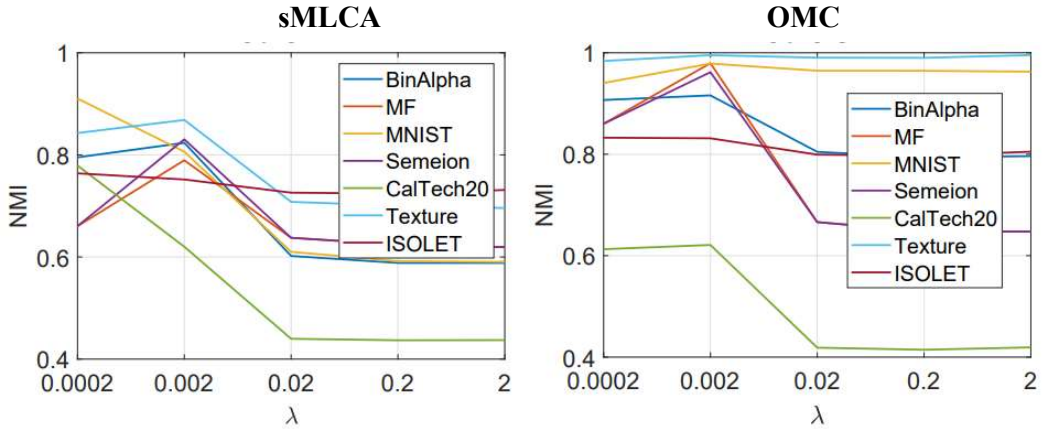

Figure 3: The NMI of our methods against different λ.

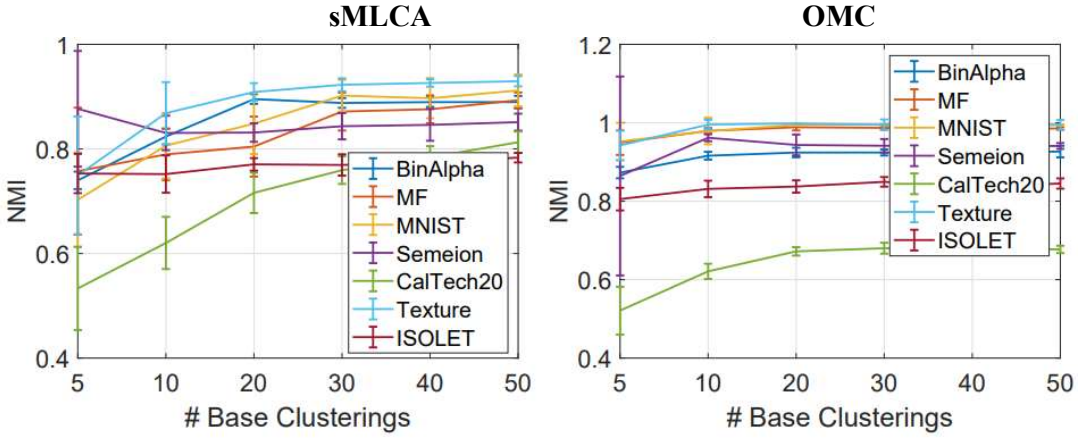

Figure 4: The NMI of our methods with different numbers of base clustering, where the vertical error bar indicates the standard deviation over 20 repetitions

Analysis of the Clustering Performance:

The clustering performance of each approach across 7 data sets is shown in Tables 1–7. These data sets contain the following observations. Firstly, virtually always outperforming all approaches in comparison, including sMLCA and OMC, the proposed methods demonstrate the applicability of the proposed model's enhanced Co-association matrix to various clustering techniques. However, OMC typically surpasses sMLCA, indicating that the refined co-association is better suited for Meta clustering. Second, the suggested methods have made some notable advances. For instance, on BinAlpha, OMC improves the ACC from 0.45 to 0.85 when

compared to the best approach in the comparison. The greatest ACC of the compared algorithms on CalTech20 is 0.49, whereas ACC sMLCA's is 0.72. Significant improvements have been made by the suggested methods in terms of various measures. Additionally, OMC does incredibly well on MF, MNIST, Semeion, and Texture; all metrics are very close to 1. These events imply that the suggested model results in a clustering ensemble success. Third, the suggested model's highly competitive performance is attained with **A** set hyper-parameter, demonstrating the model's viability. Additionally, because sMLCA and OMC consistently generate higher clustering performance on all the data sets, the proposed model is robust to various data sets.

BinAlpha	TA-CL	TA-SL	PTGP	LWSC	LWGP	E-HC	E-MC	DREC	SPCE	SEC	sMLCA	OMC
ACC	0.42	0.18	0.42	0.42	0.43	0.37	0.45	0.37	0.29	0.44	0.71	**0.85**
NMI	0.57	0.30	0.57	0.57	0.57	0.53	0.59	0.51	0.54	0.58	0.82	**0.91**
Purity	0.45	0.19	0.44	0.44	0.45	0.38	0.47	0.39	0.28	0.47	0.71	**0.87**
ARI	0.29	0.08	0.29	0.28	0.28	0.26	0.30	0.24	0.22	0.29	0.64	**0.81**
F1-score	0.31	0.12	0.31	0.30	0.30	0.29	0.32	0.27	0.30	0.31	0.65	**0.82**
Precision	0.27	0.07	0.27	0.27	0.27	0.22	0.30	0.23	0.29	0.29	0.55	**0.80**
Recall	0.36	0.63	0.36	0.34	0.34	0.45	0.33	0.32	0.31	0.32	0.79	**0.84**

Table 1: Clustering Performance on BinAlpha (# samples: 1404, dimension: 320, # clusters: 36)

MF	TA-CL	TA-SL	PTGP	LWSC	LWGP	E-HC	E-MC	DREC	SPCE	SEC	sMLCA	OMC
ACC	0.60	0.50	0.64	0.67	0.64	0.58	0.65	0.36	0.58	0.59	0.71	**0.99**
NMI	0.63	0.53	0.65	0.65	0.65	0.61	0.65	0.34	0.62	0.60	0.79	**0.97**
Purity	0.64	0.53	0.67	0.69	0.67	0.61	0.67	0.38	0.61	0.62	0.71	**0.99**
ARI	0.50	0.37	0.52	0.53	0.53	0.48	0.52	0.25	0.45	0.47	0.68	**0.97**
F1-score	0.55	0.45	0.57	0.58	0.58	0.54	0.57	0.37	0.52	0.52	0.72	**0.98**
Precision	0.51	0.34	0.53	0.55	0.53	0.47	0.54	0.31	0.42	0.49	0.58	**0.98**
Recall	0.60	0.71	0.62	0.61	0.64	0.63	0.61	0.73	0.71	0.56	0.96	**0.98**

Table 2: Clustering Performance on MF (# samples: 2000, dimension: 649, # clusters: 10)

MNIST	TA-CL	TA-SL	PTGP	LWSC	LWGP	E-HC	E-MC	DREC	SPCE	SEC	sMLCA	OMC
ACC	0.65	0.20	0.66	0.61	0.57	0.60	0.65	0.48	0.54	0.53	0.79	**0.97**
NMI	0.61	0.13	0.62	0.61	0.59	0.60	0.63	0.43	0.48	0.52	0.80	**0.97**
Purity	0.66	0.20	0.68	0.66	0.62	0.62	0.69	0.49	0.55	0.58	0.79	**0.98**
ARI	0.50	0.05	0.52	0.48	0.46	0.49	0.52	0.34	0.42	0.38	0.73	**0.96**
F1-	0.55	0.21	0.57	0.54	0.52	0.55	0.57	0.42	0.44	0.45	0.76	**0.97**

score												
Precision	0.52	0.12	0.54	0.49	0.45	0.44	0.54	0.37	0.31	0.42	0.66	**0.96**
Recall	0.59	0.95	0.60	0.60	0.60	0.74	0.61	0.57	0.83	0.48	0.91	**0.97**

Table 3: Clustering Performance on MNIST (# samples: 5000, dimension: 784, # clusters: 10)

Semeion	TA-CL	TA-SL	PTGP	LWSC	LWGP	E-HC	E-MC	DREC	SPCE	SEC	sMLCA	OMC
ACC	0.70	0.42	0.69	0.68	0.62	0.63	0.67	0.45	0.57	0.59	0.84	**0.98**
NMI	0.63	0.41	0.63	0.63	0.59	0.60	0.63	0.38	0.57	0.56	0.83	**0.96**
Purity	0.70	0.44	0.70	0.70	0.65	0.64	0.70	0.46	0.60	0.63	0.84	**0.98**
ARI	0.51	0.24	0.50	0.50	0.46	0.48	0.50	0.29	0.40	0.41	0.79	**0.96**
F1-score	0.56	0.36	0.56	0.55	0.52	0.54	0.56	0.39	0.47	0.48	0.81	**0.96**
Precision	0.52	0.24	0.52	0.52	0.46	0.46	0.52	0.32	0.38	0.44	0.74	**0.96**
Recall	0.61	0.71	0.60	0.60	0.60	0.64	0.59	0.66	0.66	0.52	0.89	**0.96**

Table 4: Clustering Performance on Semeion (# samples: 1593, dimension: 256, # clusters: 10)

CalTech20	TA-CL	TA-SL	PTGP	LWSC	LWGP	E-HC	E-MC	DREC	SPCE	SEC	sMLCA	OMC
ACC	0.34	0.42	0.34	0.32	0.33	0.45	0.36	0.34	0.49	0.29	**0.72**	0.41
NMI	0.40	0.26	0.40	0.39	0.40	0.45	0.42	0.35	0.45	0.38	0.62	**0.62**
Purity	0.63	0.52	0.63	0.64	0.64	0.64	0.66	0.59	0.66	0.63	0.73	**0.78**
ARI	0.26	0.18	0.26	0.22	0.22	0.35	0.25	0.22	0.39	0.20	**0.78**	0.32
F1-score	0.33	0.36	0.33	0.29	0.29	0.43	0.33	0.31	0.45	0.26	**0.82**	0.38
Precision	0.56	0.28	0.56	0.52	0.51	0.53	0.54	0.47	0.50	0.52	**0.76**	0.74
Recall	0.24	0.56	0.24	0.20	0.21	0.37	0.23	0.25	0.44	0.18	**0.89**	0.25

Table 5: Clustering Performance on CalTech20 (# samples: 2386, dimension: 30,000, # clusters: 20)

Texture	TA-CL	TA-SL	PTGP	LWSC	LWGP	E-HC	E-MC	DREC	SPCE	SEC	sMLCA	OMC
ACC	0.71	0.41	0.73	0.71	0.68	0.67	0.67	0.41	0.63	0.61	0.86	**0.99**
NMI	0.72	0.43	0.73	0.74	0.73	0.70	0.71	0.41	0.69	0.63	0.86	**0.99**
Purity	0.72	0.42	0.74	0.74	0.72	0.68	0.69	0.44	0.65	0.64	0.86	**0.99**
ARI	0.60	0.23	0.61	0.62	0.60	0.56	0.58	0.29	0.53	0.48	0.81	**0.99**
F1-	0.63	0.35	0.65	0.66	0.64	0.61	0.62	0.39	0.59	0.53	0.83	**0.99**

score												
Precision	0.59	0.22	0.62	0.63	0.59	0.54	0.58	0.33	0.46	0.50	0.78	**0.99**
Recall	0.69	0.89	0.68	0.70	0.71	0.71	0.67	0.75	0.82	0.58	0.89	**0.99**

Table 6: Clustering Performance on Texture (# samples: 5500, dimension: 20, # clusters: 11)

ISOLET	TA-CL	TA-SL	PTGP	LWSC	LWGP	E-HC	E-MC	DREC	SPCE	SEC	sMLCA	OMC
ACC	0.54	0.39	0.53	0.55	0.52	0.45	0.58	0.32	0.57	0.55	0.57	**0.67**
NMI	0.71	0.58	0.71	0.72	0.71	0.66	0.74	0.41	0.81	0.71	0.75	**0.83**
Purity	0.57	0.40	0.56	0.59	0.56	0.46	0.61	0.35	0.30	0.59	0.58	**0.70**
ARI	0.49	0.31	0.49	0.48	0.47	0.44	0.51	0.25	0.36	0.48	0.56	**0.63**
F1-score	0.52	0.35	0.51	0.50	0.49	0.47	0.53	0.30	0.38	0.50	0.58	**0.65**
Precision	0.46	0.23	0.46	0.45	0.43	0.35	0.49	0.25	0.58	0.46	0.48	**0.62**
Recall	0.59	**0.78**	0.59	0.56	0.57	0.74	0.58	0.69	0.32	0.55	0.75	0.68

Table 7: Clustering Performance on ISOLET (# samples: 7791, dimension: 617, # clusters: 26)

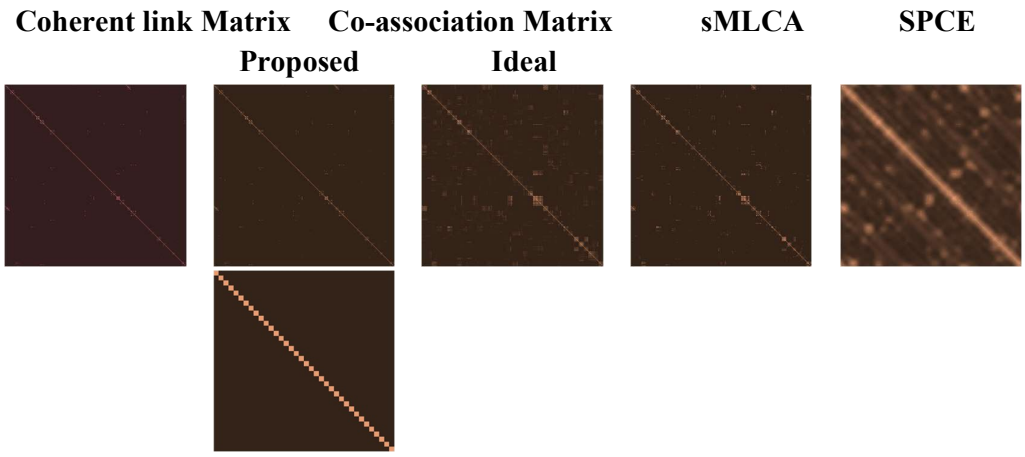

Figure 5: Visual comparison of the learned pairwise similarity matrices for different

methods. All the matrices share the samecolor bar, and the brighter color indicates a larger value.

Comparison against Base Clustering:

The average NMI of our techniques was compared to the average NMI of each base cluster in the candidate clustering pool in Fig. 2. It is notable that both OMC and sMLCA can greatly increase the NMI of the base clustering on all of the data sets, with OMC usually superior sMLCA.

Sensitivity to Hyper-parameter:

The NMI of the proposed methods with different values of is shown in Fig. 3, from which we can draw the following conclusions: first, for most data sets, the highest NMI occurs when = 0.002 for both sMLCA and OMC, demonstrating the importance of removing the incorrect links from the initial co-association matrix; and second, a smaller typically leads to better clustering performance for both sMLCA and OMC.

Performance with Different Number of Base clustering:

The observations are shown in Fig. 4, which shows how the proposed model is affected by various base clustering numbers. First, base clustering improves clustering performance as seen by the fact that the NMIs of both sMLCA and OMC generally rise with base clustering numbers. Second, more base clustering generally results in fewer standard deviations for all the data sets, which shows that base clustering can improve the consistency of our methods. Third, the bulk of data sets only require 20 base clustering for our algorithms to provide high NMI values.

Comparison of the Learned Pairwise Similarity Matrix:

The ideal similarity matrix of BinAlpha is shown in Fig. 5, along with the cohere-link matrix, the conventional co-association matrix, and the learned co-association matrices by MLCA (Huang, Wang, and Lai 2018), SPCE (Zhou 2020), and the proposed model. All of the matrices are normalized to [0, 1], and they all have the same colour bar. Form 5 shows that while the Co-association matrix is dense and contains many wrong connections, the Cohesive Matrix is sparse and has the majority of its connections right. The improved Co-association matrix of the suggested model is very near to the ideal one because it takes use of the label correspondence of the 3-D matrix created by the stacking of the Cohesive Matrix and the association matrix. High clustering performance was achieved because, despite minor mistake corrections, nearly all the relationships between two samples belonging to the same cluster had been accurately retrieved. In contrast, both the affinity matrices of LWGP and SPCE contain a large number of wrong connections, but not enough correct connections, which explains why they performed worse at clustering than the suggested model.

Conclusion

We initially introduced the clustering ensemble to matric of similar label Matrix approximation. In contrast to earlier approaches, the proposed model addresses clustering ensemble from a global perspective, i.e., by taking advantage of the 3-D matrix formed by the Cohesive Matrix and the co-association matrix's, which allows the Cohesive Matrix's useful information to be efficiently transmitted to the co-association matrix. Numerous tests have demonstrated that the proposed model:

i. It raises the clustering ensemble's current state-of-the-art performance to a new level;
ii. It is robust to different data sets in terms of the recommended value for the hyper-parameter;
iii. Only a small number of base clustering are necessary to produce high clustering performance.

References

1. Sagi, O.; and Rokach, L. 2018. Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.
2. Song, L.; Tekin, C.; and Van Der Scar, M. 2014. Online Learning in Large-scale Contextual Recommender Systems. IEEE Transactions on Services Computing.
3. Tao, Z.; Liu, H.; and Fu, Y. 2017. Simultaneous Clustering and Ensemble. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence.
4. Tao, Z.; Liu, H.; Li, S.; Ding, Z.; and Fu, Y. 2017. From ensemble clustering to multi-view clustering. In Proceedings of the 26th International Joint Conference on Artificial Intelligence.
5. Tao, Z.; Liu, H.; Li, S.; and Fu, Y. 2016. Robust Spectral Ensemble Clustering. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management.
6. Ulrike, L. 2007. A Tutorial on Spectral Clustering. Stat. Computer.
7. Vega-Pons, S.; and Ruiz-Shulcloper, J. 2011. A Survey of Clustering Ensemble Algorithms. International Journal of Pattern Recognition and Artificial Intelligence.
8. Wu, W.; Jia, Y.; Kwong, S.; and Hou, J. 2018. Pairwise Constraint Propagation-Induced Symmetric Nonnegative Matrix Factorization. IEEE Transactions on Neural Networks and Learning Systems.
9. Wu, W.; Kwong, S.; Zhou, Y.; Jia, Y.; and Gao, W. 2018. Nonnegative Matrix Factorization with Mixed Hypergraph Regularization for Community Detection. Information Sciences.
10. Zhang, C.; Fu, H.; Hu, Q.; Cao, X.; Xie, Y.; Tao, D.; and Xu,D. 2020. Generalized Latent Multi-View Subspace Cluster- ing. IEEE Transactions on Pattern Analysis and Machine Intelligence.
11. Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; and Kilmer, M. 2014. Novel Methods for Multi linear Data Completion and De- noising based on Tensor-SVD. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
12. Zhou, J.; Zheng, H.; and Pan, L. 2019. Ensemble Clustering based on Dense Representation.
13. Zhou, P.; Du, L.; Liu, X.; Shen, Y.; Fan, M.; and Li, X. 2020. Self-Paced Clustering Ensemble. IEEE Transactions on Neural Networks and Learning Systems.
14. Boongoen, and Iam-On, N. 2018. Cluster Ensembles: A Survey of Approaches with Recent Extensions and Applications. Computer Science Review.
15. Fred, A. L.; and Jain, A. K. 2005. Combining Multiple Clustering's Using Evidence Accumulation. IEEE Transactions on Pattern Analysis and Machine Intelligence.

16. Gao, J.; Yamada, M.; Kaski, S.; Mamitsuka, H.; Zhu, S.; and Kambhampati, S. 2016. A Robust Convex Formulations for Ensemble Clustering. In Proceedings of the 25th Inter- national Joint Conference on Artificial Intelligence (IJCAI 2016), AAAI Press International Joint Conferences on Artificial Intelligence.
17. Huang, D.; Lai, J.-H. And Wang, 2016. Robust Ensemble Clustering Using Probability Trajectories. IEEE Transactions on knowledge and Data Engineering.
18. Huang, D.; Wang, C.; Peng, H.; Lai, J.; and Kwoh, C. 2018. Enhanced Ensemble Clustering via Fast Propagation of Cluster-Wise Similarities. IEEE Transactions on Systems, Man, and Cybernetics: Systems.
19. Huang, D.; Wang, C.-D.; and Lai, J.-H. 2018. Locally Weighted Ensemble Clustering. IEEE Transactions on Cybernetics.
20. Jain, A. K. 2010. Data Clustering: 50 Years beyond K- means. Pattern Recognition Letters.
21. Jia, Y.; Hou, J.; and Kwong, S. 2020. Constrained Clustering With Dissimilarity Propagation-Guided Graph-Laplacian PCA. IEEE Transactions on Neural Networks and Learning Systems.
22. Jia, Y.; Kwong, S.; and Hou, J. 2018. Semi-Supervised Spectral Clustering with Structured sparsely Regularization. IEEE Signal Processing Letters.
23. Jia, Y.; Kwong, S.; Hou, J.; and Wu, W. 2018. Convex Constrained Clustering with Graph-Laplacian. In 2018 IEEE International Conference on Multimedia and Expo (ICME).
24. Jia, Y.; Kwong, S.; Hou, J.; and Wu, W. 2020a. Semi- supervised Non-Negative Matrix Factorization with Dis- similarity and Similarity Regularization. IEEE Transactions on Neural Networks and Learning Systems.
25. Jia, Y.; Liu, H.; Hou, J.; and Kwong, S. 2020b. Clustering- Aware Graph Construction: A Joint Learning Perspective. IEEE Transactions on Signal and Information Processing over Networks.
26. Jia, Y.; Liu, H.; Hou, J.; and Kwong, S. 2020c. Semi supervised Adaptive Symmetric Non-Negative Matrix Factorization. IEEE Transactions on Cybernetics.
27. Jia, Y.; Wu, W.; Wang, R.; Hou, J. and Kwong, S. 2020d. Joint Optimization for Pairwise Constraint Propagation. IEEE Transactions on Neural Networks and Learning Systems.
28. Johnson, S. C. 1967. Hierarchical Clustering Schemes.
29. Kilmer, M. E.; Braman, K.; Hao, N.; and Hoover, R. C. 2013. Third-order Tensors as Operators on Matrices: A Theoretical and Computational Framework with Applications in Imaging. SIAM Journal on Matrix Analysis and Applications.
30. Li, H.; Kwong, S.; Chen, C.; Jia, Y.; and Cong, R. 2019. Super pixel Segmentation based on Square-wise Asymmetric Partition and Structural Approximation. IEEE Transactions on Multimedia Liu, H.; Jia, Y.; Hou, J.; and Zhang, Q. 2019. Imbalance aware Pairwise Constraint Propagation. In In Proceedings of the 27th ACM International Conference on Multimedia.

31. Neuroquantology | December 2022 | Volume 20 | Issue 22 |PAGE 2768-2774|DOI: 10.48047/NQ.2022.20.22.NQ10269 MS Urvashi Soni et al/A Study of Consensus Function in Cluster Ensemble
32. Liu, H.; Wu, J.; Liu, T.; Tao, D.; and Fu, Y. 2017. Spectral Ensemble Clustering via Weighted K-Means: Theoretical and Practical Evidence. IEEE Transactions on Knowledge and Data Engineering.
33. Moore, A. W. 1999. Very Fast EM-based Mixture Model Clustering Using Multi resolution trees. In Advances in Neural information processing systems.
34. Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2002. On Spectral Clustering: Analysis and an Algorithm. In Advances in Neural Information Processing Systems.