Projective Clustering Ensembles

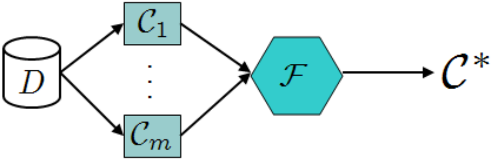
F. Gullo * C. Domeniconi † A. Tagarelli *

* Dept. of Electronics, Computer and Systems Science
University of Calabria, Italy

† Dept. of Computer Science
George Mason University, Virginia (USA)

IEEE International Conference on Data Mining (ICDM) *2009*

Clustering Ensembles

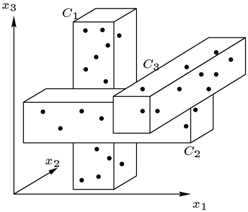


input a set $\mathcal{E} = \{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ of clustering solutions (i.e., *ensemble*)

output a *consensus partition* $\mathcal{C}^*$ computed according to a *consensus function* $\mathcal{F}$

goal : to reduce the (inevitable) bias of any clustering solution due to the peculiarities of the specific clustering algorithm being used (*ill-posed* nature of clustering)

Projective Clustering



input a set $\mathcal{D}$ of D-dimensional points (data objects)

output a partition $\mathcal{C}$ of $\mathcal{D}$, a set $\mathcal{S}$ of *subspaces* s.t. each $S \in \mathcal{S}$ is assigned to one (and only one) cluster $C \in \mathcal{C}$

- *goal* : overcoming issues due to *the curse of dimensionality*
- *assumption* : objects within the same cluster C are close to each other if (and only if) they are projected onto the subspace S associated to C

figure borrowed from [Procopiuc et Al., SIGMOD'02]

Clustering Ensembles and Projective Clustering have been so far considered as two distinct problems...

Projective Clustering Ensembles (PCE)

PCE problem addressed for the first time:

given a set of *projective clustering solutions* (i.e., a *projective ensemble*), the objective is to discover a *projective consensus partition*

Challenge:

information about *feature-to-cluster* assignments have to be considered: traditional clustering ensembles methods do not work!

Contributions

- rigorous formulations of PCE as an optimization problem
  - *two-objective PCE*
  - *single-objective PCE*
- well-founded heuristics for each formulation
  - *MOEA-PCE*
  - *EM-PCE*

Outline

Projective clustering solution

Definition (projective clustering solution)

Let $\mathcal{D} = \{\vec{o}_1, \ldots, \vec{o}_N\}$ be a set of D-dimensional points (data objects). A *projective clustering solution* C defined over $\mathcal{D}$ is a triple $\langle \mathcal{L}, \Gamma, \Delta \rangle$:

- $\mathcal{L} = \{\ell_1, \ldots, \ell_K\}$ is a set of cluster labels which uniquely represent the K clusters
- $\Gamma : \mathcal{L} \times \mathcal{D} \to S_\Gamma$ is a function which stores the probability that object $\vec{o}_n$ belongs to the cluster labeled with ℓ_k, $\forall k \in [1..K]$, $n \in [1..N]$, such that $\sum_{k=1}^{K} \Gamma_{kn} = 1, \forall n \in [1..N]$, where Γ_{kn} hereinafter refers to $\Gamma(\ell_k, \vec{o}_n)$
- $\Delta : \mathcal{L} \times [1..D] \to [0, 1]$ is a function which stores the probability that the d-th feature is a relevant dimension for the objects in the cluster labeled with ℓ_k, $\forall k \in [1..K]$, $d \in [1..D]$, such that $\sum_{d=1}^{D} \Delta_{kd} = 1, \forall k \in [1..K]$, where Δ_{kd} hereinafter refers to $\Delta(\ell_k, d)$

Two-objective PCE

Motivation:

A projective consensus partition $C^* = \langle \mathcal{L}^*, \Gamma^*, \Delta^* \rangle$ derived from an ensemble $\mathcal{E}$ should meet requirements related to:

- the data object clustering of the solutions in $\mathcal{E}$
- the feature-to-cluster assignment of the solutions in $\mathcal{E}$

$\Longrightarrow$ PCE can be naturally formulated considering two objectives

Two-objective PCE: formulation

$$C^* = \arg\min_{\hat{C}} \left[\Psi_o(\hat{C}, \mathcal{E}, \mathcal{D}),\ \Psi_f(\hat{C}, \mathcal{E}, \mathcal{D}) \right]$$

where

$$\Psi_o(\hat{C}, \mathcal{E}, \mathcal{D}) = \sum_{C \in \mathcal{E}} \frac{1}{2} \Big(\psi_o(\hat{C}, C) + \psi_o(C, \hat{C}) \Big)$$

$$\Psi_f(\hat{C}, \mathcal{E}, \mathcal{D}) = \sum_{C \in \mathcal{E}} \frac{1}{2} \Big(\psi_f(\hat{C}, C) + \psi_f(C, \hat{C}) \Big)$$

and $\psi_o(C_i, C_j)$ (resp. $\psi_f(C_i, C_j)$) is computed by resorting to the extended Jaccard similarity coefficient applied to the Γ_{kn} (resp. Δ_{kd}) values of C_i and C_j

Two-objective PCE: heuristic

- two-objective PCE formulation: objectives are conflicting with each other
- naïve solutions given by (linear) combining the two objectives into a single one have several drawbacks:
  - mixing non-commensurable objectives
  - hard setting of the weights needed for the linear combination
  - prior knowledge of the application domain
- *idea*: resort to the *Multi Objective Evolutionary Algorithms* (MOEAs) domain
  $\Longrightarrow$ we exploit *NSGA-II* algorithm

Two-objective PCE: MOEA-PCE algorithm

MOEA-PCE Algorithm

Input: a projective ensemble $\mathcal{E}$ of size M, defined over a set $\mathcal{D}$ of N D-dimens. objects; the number K of clusters in the output projective consensus partitions; the population size t; the max number I of iterations

Output: a set $\mathcal{S}^*$ of projective consensus partitions

1: $\mathcal{S} \leftarrow populationRandomGen(\mathcal{E}, t, K)$, $it \leftarrow 1$
2: **repeat**
3: $\rho \leftarrow computeParetoRanking(\mathcal{S})$
4: $\langle \mathcal{S}', \mathcal{S}'' \rangle \leftarrow \langle \check{\mathcal{S}}' \subset \mathcal{S},\ \check{\mathcal{S}}'' \subset \mathcal{S} \rangle\ :\ |\check{\mathcal{S}}'| = |\mathcal{S}|/2,\ |\check{\mathcal{S}}''| = |\mathcal{S}|/2,\ \check{\mathcal{S}}' \cup \check{\mathcal{S}}'' = \mathcal{S},\ \rho(x') \leq \rho(x''), \forall x' \in \check{\mathcal{S}}', x'' \in \check{\mathcal{S}}''$
5: $\mathcal{S}'_{CM} \leftarrow crossoverAndMutation(\mathcal{S}')$
6: $\mathcal{S} \leftarrow \mathcal{S}' \cup \mathcal{S}'_{CM}$
7: $it \leftarrow it + 1$
8: **until** $it = I$
9: $\rho \leftarrow computeParetoRanking(\mathcal{S})$
10: $\mathcal{S}^* \leftarrow \{x' \in \mathcal{S}\ :\ \rho(x') \leq \rho(x''), \forall x'' \in \mathcal{S}, x'' \neq x'\}$

Two-objective PCE: MOEA-PCE algorithm (2)

- The proposed MOEA-PCE heuristic is based on the classic MOEA notions of:
  - *domination*
  - *Pareto-optimality*
  - *Pareto-ranking* function (ρ)
- MOEA-PCE works in $\mathcal{O}(I\ t\ M\ K^2\ (N+D))$

Two-objective PCE: MOEA-PCE algorithm (3)

Weaknesses of MOEA-PCE:

- high complexity in the approach
- efficiency (mostly due to I)
- hard setting for I and t
- results not easily interpretable (multiple output results)

Single-objective PCE: formulation

PCE formulation alternative to two-objective PCE:

$$
\begin{aligned}
C^* \quad &= \quad \arg\min_{\hat{C}} Q(\hat{C}, \mathcal{E}) \\
s.t.& \\
& \sum_{k=1}^{K} \hat{\Gamma}_{kn} = 1, \quad \forall n \in [1..N] \\
& \sum_{d=1}^{D} \hat{\Delta}_{kd} = 1, \quad \forall k \in [1..K] \\
& \hat{\Gamma}_{kn} \geq 0,\ \hat{\Delta}_{kd} \geq 0, \quad \forall k \in [1..K], n \in [1..N], d \in [1..D]
\end{aligned}
$$

where

$$
Q(\hat{C}, \mathcal{E}) = \sum_{k=1}^{K} \sum_{n=1}^{N} \hat{\Gamma}_{kn}^{\alpha} \sum_{h=1}^{H} \gamma_{hn} \sum_{d=1}^{D} \left(\hat{\Delta}_{kd} - \delta_{hd} \right)^2
$$

Single-objective PCE: formulation (2)

$$Q(\hat{C}, \mathcal{E}) = \sum_{k=1}^{K} \sum_{n=1}^{N} \hat{\Gamma}_{kn}^{\alpha} \sum_{h=1}^{H} \gamma_{hn} \sum_{d=1}^{D} \left(\hat{\Delta}_{kd} - \delta_{hd}\right)^2$$

Rationale of function Q at the basis of the proposed single-objective PCE formulation:

- it embeds both object-based and feature-based representations of the solutions in the ensemble
- it is essentially based on measuring, for each object, the "distance error" between the feature-based representation of the clusters in the consensus partition and the clusters in the solutions of the ensemble
- the discrepancy between two clusters is weighted by the probability that the object belongs to both (i.e., $\Gamma_{kn} \times \gamma_{hn}$)

Single-objective PCE: heuristic

A procedure inspired to the popular EM has been defined

Unconstrained function Q_λ is derived by applying Lagrangian multipliers:

$$Q_\lambda(\hat{C},\mathcal{E}) = Q(\hat{C},\mathcal{E}) + \sum_{n=1}^{N} \lambda'_n \left(\sum_{k'=1}^{K} \hat{\Gamma}_{k' n} - 1 \right) + \sum_{k=1}^{K} \lambda''_k \left(\sum_{d'=1}^{D} \hat{\Delta}_{kd'} - 1 \right)$$

Two systems of equations are solved to derive optimal Γ^*_{kn} and Δ^*_{kd} values:

$$\Gamma^*_{kn} = \begin{cases} \dfrac{\partial\ Q_\lambda}{\partial\ \hat{\Gamma}_{kn}} = 0 \\[2ex] \dfrac{\partial\ Q_\lambda}{\partial\ \lambda'_n} = 0 \end{cases} \qquad \Delta^*_{kd} = \begin{cases} \dfrac{\partial\ Q_\lambda}{\partial\ \hat{\Delta}_{kd}} = 0 \\[2ex] \dfrac{\partial\ Q_\lambda}{\partial\ \lambda''_k} = 0 \end{cases}$$

Single-objective PCE: heuristic (2)

The solutions of the systems of equations are:

$$\Gamma^*_{kn} = \left[\sum_{k'=1}^{K} \left(\frac{X_{kn}}{X_{k'n}} \right)^{\frac{1}{\alpha-1}} \right]^{-1} \qquad \Delta^*_{kd} = \frac{Z_{kd}}{Y_k}$$

where

$$X_{kn} = \sum_{h=1}^{H} \gamma_{hn} \sum_{d=1}^{D} \left(\hat{\Delta}_{kd} - \delta_{hd} \right)^2$$

$$Y_k = \sum_{n=1}^{N} \hat{\Gamma}^{\alpha}_{kn} \sum_{h=1}^{H} \gamma_{hn}$$

$$Z_{kd} = \sum_{n=1}^{N} \hat{\Gamma}^{\alpha}_{kn} \sum_{h=1}^{H} \gamma_{hn} \; \delta_{hd}$$

Single-objective PCE: EM-PCE algorithm

EM-PCE Algorithm

Input: a projective ensemble $\mathcal{E}$ defined over a set $\mathcal{D}$ of data objects; the number K of clusters in the output projective consensus partition;
Output: the projective consensus partition C^*

1: $\mathcal{L}^* \leftarrow \{1, \ldots, K\}$
2: $\langle \Gamma^*, \Delta^* \rangle \leftarrow \mathit{randomGen}(\mathcal{E}, K)$
3: **repeat**
4: compute Γ^*_{kn} values
5: compute Δ^*_{kd} values
6: **until** *convergence*
7: $C^* = \langle \mathcal{L}^*, \Gamma^*, \Delta^* \rangle$

- EM-PCE converges to a local optimum of function Q
- EM-PCE works in $\mathcal{O}(I\ M\ K^2\ N\ D)$

Evaluation methodology: datasets

- eight benchmark datasets from the UCI Machine Learning Repository (Iris, Wine, Glass, Ecoli, Yeast, Segmentation, Abalone, Letter)
- two time-series datasets from the UCR Time Series Classification/Clustering Page (Tracedata, ControlChart)

dataset	*objects*	*attributes*	*classes*
Iris	150	4	3
Wine	178	13	3
Glass	214	10	6
Ecoli	327	7	5
Yeast	1,484	8	10
Segmentation	2,310	19	7
Abalone	4,124	7	17
Letter	7,648	16	10
Tracedata	200	275	4
ControlChart	600	60	6

Evaluation methodology: assessment criteria

Accuracy of output consensus partitions $\check{\mathcal{C}} = \langle \check{\mathcal{L}}, \check{\Gamma}, \check{\Delta} \rangle$, $|\check{\mathcal{L}}| = \check{K}$, was evaluated in terms of:

- *similarity w.r.t. (hard) reference classification* $\widetilde{\mathcal{C}}$
  - object-based representation
  - feature-based representation
- *error-rate* E [Domeniconi et Al., SDM'04] (internal criterion):

$$E(\check{\mathcal{C}}) = \sum_{k=1}^{\check{K}} \sum_{d=1}^{D} \check{\Delta}_{kd} \left(\sum_{n=1}^{N} \check{\Gamma}_{kn} \right)^{-1} \sum_{n=1}^{N} \check{\Gamma}_{kn} \left(\overline{c}_{kd} - o_{nd} \right)^2$$

Results: evaluation w.r.t. reference classification

Object-based representation

	ensemble	*MOEA-PCE*			*EM-PCE*		
data	*avg-max*	*avg*	*max-std*	*gain w.r.t. ens. (avg)*	*avg*	*max-std*	*gain w.r.t. ens. (avg)*
Iris	.632 .925	.919	.925 .015	+.287	.762	.767 .040	+.130
Wine	.738 .910	.913	.928 .105	+.175	.782	.840 .028	+.044
Glass	.565 .775	.683	.768 .046	+.118	.639	.644 .002	+.074
Ecoli	.421 .689	.603	.686 .054	+.182	.329	.419 .040	-.092
Yeast	.675 .750	.723	.745 .015	+.048	.638	.641 .001	-.037
Segm.	.590 .821	.755	.835 .049	+.165	.653	.663 .004	+.063
Abal.	.509 .520	.518	.558 .043	+.009	.512	.542 .002	+.003
Letter	.522 .640	.597	.612 .031	+.075	.554	.562 .006	+.032
Trace	.772 .868	.862	.998 .059	+.090	.875	.935 .030	+.103
Contr.	.681 .981	.895	.965 .049	+.214	.790	.806 .007	+.109

Results: evaluation w.r.t. reference classification (2)

Object-based representation

- both MOEA-PCE and EM-PCE achieved accuracy comparable or far better than that reached on average by the solutions in the ensemble
- avg gains: +13.6% (MOEA-PCE) and +4.3% (EM-PCE)
- max gains: +29% (MOEA-PCE, on Iris) and +13% (EM-PCE, on Iris)

Results: evaluation w.r.t. reference classification (3)

Feature-based representation

	ensemble	*MOEA-PCE*			*EM-PCE*		
data	*avg-max*	*avg*	*max-std*	*gain w.r.t. ens. (avg)*	*avg*	*max-std*	*gain w.r.t. ens. (avg)*
Iris	.662 .998	.988	1 .029	+.326	.845	.895 .043	+.183
Wine	.822 .989	.955	.997 .027	+.133	.869	.899 .080	+.047
Glass	.731 .891	.851	.900 .027	+.120	.817	.877 .041	+.086
Ecoli	.763 .879	.858	.884 .016	+.095	.903	.953 .052	+.140
Yeast	.720 .805	.790	.804 .009	+.070	.684	.690 .003	-.036
Segm.	.618 .720	.729	.737 .049	+.111	.625	.632 .008	+.007
Abal.	.716 .754	.759	.849 .023	+.043	.726	.748 .013	+.010
Letter	.646 .693	.767	.818 .012	+.121	.780	.786 .007	+.134
Trace	.661 .818	.755	.811 .0.25	+.094	.753	.773 .021	+.092
Contr.	.663 .894	.880	.910 .016	+.217	.734	.774 .022	+.071

Results: evaluation w.r.t. reference classification (4)

Feature-based representation

- results comparable to the object-based representation case
- avg gains: +13.3% (MOEA-PCE) and +7.3% (EM-PCE)
- max gains: +32.6% (MOEA-PCE, on Iris) and +18.3% (EM-PCE, on Iris)

Results: evaluation in terms of error rate

- both MOEA-PCE and EM-PCE outperformed average results by the solutions in the ensemble and by reference classification
- avg gains w.r.t. ensemble: $+0.358\%$ (MOEA-PCE) and $+0.27\%$ (EM-PCE)
- avg gains w.r.t. reference classification: $+0.6\%$ (MOEA-PCE) and $+0.51\%$ (EM-PCE)

Conclusion

- Projective Clustering Ensembles (PCE) problem addressed for the first time
- Two formulations of PCE as an optimization problem
  - Two-objective PCE
  - Single-objective PCE
- Heuristic algorithms for each one of the proposed formulations
  - MOEA-PCE
  - EM-PCE
- Accuracy improvements achieved by both the proposed heuristics w.r.t. avg ensemble results in terms of external as well as internal criteria

Thanks!