Adaptive batch SOM for multiple dissimilarity data tables

Anderson Dantas
absd@cin.ufpe.br

Introduction

- Clustering methods organize a set of items into clusters
- Items within a given cluster have a high degree of similarity
- Feature data and Relational data

Objective

Hard clustering algorithm that is able to partition objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices

Objective

- Partition
- Prototype for each cluster
- **Relevance weight**

Adaptive batch SOM for data based on multiple dissimilarity matrices

Iterative three-step algorithm
Representation, weighting and affectation

Cost function

$$J = \sum_{e_i \in E} \sum_{l=1}^{c} K^T(\delta(\chi(e_i), l)) D_{\boldsymbol{\lambda}_l}(e_i, G_l)$$

- $D_{\boldsymbol{\lambda}_l}$ is the global matching
- relevance weight vector $\boldsymbol{\lambda}_l = (\lambda_{l1}, \ldots, \lambda_{lp})$ of the dissimilarity matrices $\mathbf{D}_j$
- Neighbourhood through kernel functions K

Matching function

$$D_{\boldsymbol{\lambda}_l}(e_i, G_l) = \sum_{j=1}^{p} \lambda_{lj} D_j(e_i, G_l) = \sum_{j=1}^{p} \lambda_{lj} d_j(e_i, G_l)$$

Representation step: computation of the best prototypes

Compute the prototype $G_l^{(t)} = G^* \in E^{(q)}$ of cluster $P_l^{(t-1)}$ ($l = 1, \ldots, c$) according to:

$$G^* = argmin_{e \in E^{(q)}} \sum_{e_i \in E} K^T(\delta(\chi^{(t-1)}(e_i), l)) \sum_{j=1}^{p} \lambda_{lj}^{(t-1)} d_j(e_i, G_l)$$

Weighting step: definition of the best vectors of weights

The vectors of weights $\boldsymbol{\lambda}_l{}^{(t)} = (\lambda_{l1}^{(t)}, \ldots, \lambda_{lp}^{(t)})$ $(l = 1, \ldots, c)$, under $\lambda_{lj}^{(t)} > 0$ and $\prod_{j=1}^{p} \lambda_{lj}^{(t)} = 1$, have their weights $\lambda_{lj}^{(t)}$ $(j = 1, \ldots, p)$ calculated according to:

$$\lambda_{lj}^{(t)} = \frac{\left\{ \prod_{h=1}^{p} \left[\sum_{e_i \in E} K^T(\delta(\chi^{(t-1)}(e_i), l)) d_h(e_i, G_l) \right] \right\}^{\frac{1}{p}}}{\left[\sum_{e_i \in E} K^T(\delta(\chi^{(t-1)}(e_i), l)) d_j(e_i, G_l) \right]}$$

Affectation step: definition of the best partition

$$m = (\chi^{(t)}(e_i))^{(t)} = argmin_{1 \leq r \leq c} \sum_{l=1}^{c} K^T(\delta(r,l)) \sum_{j=1}^{p} \lambda_{lj}^{(t)} d_j(e_i, G_l)$$

Empirical results

- Databases from the UCI Machine Learning Repository
- Measures: corrected Rand index (CR), F-measure, overall error rate of classification (OERC)
- Confusion matrix
- Data sets described by a matrix of objects x real-valued attributes
- All attributes x single attibutes

Wine dataset

Parameters

- Topology: 2x5;
- T_{min}: 0.3;
- T_{max}: 3.0;
- N_{iter}: 500

Indexes	B-SOM	AB-SOM
CR	0.31	0.42
$F - measure$	0.45	0.52
$OERC$	27.00%	9.00%

Confusion matrix

Cluster/Class	1	2	3	Majority Class
0,0	0	5	16	3
0,1	0	1	4	3
0,2	0	21	0	2
0,3	2	22	0	2
0,4	6	10	0	2
1,0	0	0	15	3
1,1	0	0	13	3
1,2	0	9	0	2
1,3	24	2	0	1
1,4	27	1	0	1

Final relevance weight matrix

Cluster/Matrix	1	2	3	4	5	6	7	8	9	10	11
0,0	**1.03**	0.27	**1.31**	**1.07**	0.48	0.90	**3.05**	0.33	**1.87**	**1.25**	0.5
0,1	0.99	**1.05**	0.55	**1.09**	0.27	0.69	0.79	0.52	0.41	0.92	**1.8**
0,2	**1.28**	0.52	0.41	0.43	**1.47**	0.60	**2.42**	0.90	**1.57**	**5.17**	0.1
0,3	0.69	0.33	0.93	**2.64**	**1.32**	0.48	**2.29**	**1.22**	**0.45**	**2.39**	0.3
0,4	**0.77**	**4.19**	0.27	0.99	0.16	**1.92**	0.84	**1.96**	0.24	**3.04**	**3.7**
1,0	**1.02**	0.37	0.63	0.89	0.47	**1.68**	**2.42**	0.25	**2.79**	0.45	**1.4**
1,1	0.46	0.21	0.79	0.90	**1.85**	0.49	**5.07**	0.86	0.24	0.36	**2.2**
1,2	0.42	0.61	0.51	**1.35**	0.16	**2.93**	**3.12**	0.24	**2.27**	**2.96**	**1.3**
1,3	0.77	0.55	0.69	0.16	0.29	**1.38**	**3.13**	**2.14**	**1.12**	**1.95**	**1.0**
1,4	0.48	**6.01**	0.37	0.33	0.61	**1.64**	**1.65**	0.94	0.40	**1.50**	**1.7**

Adaptive batch SOM for multiple dissimilarity data tables

Anderson Dantas
absd@cin.ufpe.br