Unsupervised Feature Selection with Adaptive Structure Learning

KDD 2015

Liang Du and Yi-Dong Shen
Institute of Software, CAS
{duliang,ydshen}@ios.ac.cn

Background

- Feature selection techniques keep few informative features to alleviate the great challenges presented by data with high dimensionality.
- Without class label, unsupervised feature selection methods chooses those features that can reveal or maintain the underlying structure of data.

Related work

- Most existing methods use all the input features (including redundant and noisy ones) to capture the various structures of data.
- Such structures are further used to guide the search of relevant features.
- A lot of methods are developed: LapScore, SPEC, EVSC, TraceRatio, UDFS, MCFS, MRSF, SPFS, FSSL, GLSPFS, JELSR, NDFS, RUFS, CGSSL, RSFS, LLCFS etc.

Motivation

- On the one hand, one need the true structures of data to identify the informative features;
- On the other hand, one need the informative features to accurately estimate the structures;
- We have to face the *chicken-and-egg dilemma* between **structure characterization** and **feature learning**.

Basic idea

- We use the selected informative features to accurately re-estimate the underlying structures.
- We develop a unified framework to integrate these two essential sub-steps: structure characterization and feature learning.

Our method: FSASL

- Adaptive global structure learning via sparse representation

Global Structure Learning

$$\min_{\mathbf{S}} \sum_{i=1}^{n} (||\mathbf{x}_i - \mathbf{X}\mathbf{s}_i||^2 + \alpha||\mathbf{s}_i||_1) \quad \text{s.t.} \quad \mathbf{S}_{ii} = 0$$

Adaptive Global Structure Learning

$$\min_{\mathbf{S},\mathbf{W}} \sum_{i=1}^{n} ||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{X}\mathbf{s}_i||^2 + \alpha||\mathbf{S}||_1 + \gamma||\mathbf{W}||_{21}$$
$$\text{s.t.} \quad \mathbf{S}_{ii} = 0, \mathbf{W}^T\mathbf{X}\mathbf{X}^T\mathbf{W} = \mathbf{I}$$

- Adaptive local structure learning with probabilistic neighborhood

Local Structure Learning

$$\min_{\mathbf{P}} \sum_{i,j} (||\mathbf{x}_i - \mathbf{x}_j||_2^2 \mathbf{P}_{ij} + \mu \mathbf{P}_{ij}^2), \ \text{s.t.} \ \mathbf{P}\mathbf{1}_n = \mathbf{1}_n, \mathbf{P} \geq 0$$

Adaptive Local Structure Learning

$$\min_{\mathbf{P},\mathbf{W}} \sum_{i,j}^{n} (||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 \mathbf{P}_{ij} + \mu \mathbf{P}_{ij}^2) + \gamma||\mathbf{W}||_{21}$$
$$\text{s.t.} \quad \mathbf{P}\mathbf{1}_n = \mathbf{1}_n, \mathbf{P} \geq \mathbf{0}, \mathbf{W}^T\mathbf{X}\mathbf{X}^T\mathbf{W} = \mathbf{I}$$

- Unsupervised Feature Selection with Adaptive Structure Learning (FSASL)

$$\min_{\mathbf{W},\mathbf{S},\mathbf{P}} \left(||\mathbf{W}^T\mathbf{X} - \mathbf{W}^T\mathbf{X}\mathbf{S}||^2 + \alpha||\mathbf{S}||_1\right) + \beta \sum_{i,j}^{n} \left(||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||^2 \mathbf{P}_{ij} + \mu \mathbf{P}_{ij}^2\right) + \gamma||\mathbf{W}||_{21}$$
$$\text{s.t.} \quad \mathbf{S}_{ii} = 0, \mathbf{P}\mathbf{1}_n = \mathbf{1}_n, \mathbf{P} \geq \mathbf{0}, \mathbf{W}^T\mathbf{X}\mathbf{X}^T\mathbf{W} = \mathbf{I}$$

Experiments

- K-means clustering with selected features

Table 1: Aggregated clustering results measured by Accuracy (%) of the compared methods.

Data Sets	AllFea	LapScore	MCFS	LLCFS	UDFS	NDFS	SPFS	RUFS	JELSR	GLSPFS	FSASL
MFEA	68.73	51.78 ± 5.51 0.00	51.04 ± 8.13 0.00	60.38 ± 8.58 0.00	48.94 ± 3.32 0.00	67.13 ± 7.53 0.01	**68.20** **± 9.43** **0.22**	64.58 ± 7.99 0.00	67.01 ± 8.37 0.01	61.00 ± 8.70 0.00	**69.94** **± 7.19** **1.00**
USPS49	77.70	69.21 ± 8.95 0.00	53.74 ± 3.50 0.00	94.96 ± 1.44 0.03	94.05 ± 1.13 0.00	68.12 ± 8.18 0.00	83.43 ± 6.66 0.00	85.86 ± 2.58 0.00	95.16 ± 0.55 0.00	94.75 ± 0.61 0.00	**95.95** **± 0.48** **1.00**
UMIST	42.40	36.73 ± 1.18 0.00	44.46 ± 3.26 0.00	47.31 ± 0.83 0.00	48.04 ± 1.92 0.00	52.80 ± 2.26 0.00	46.72 ± 1.70 0.00	50.87 ± 1.95 0.00	53.52 ± 1.54 0.01	50.53 ± 0.59 0.00	**54.92** **± 1.89** **1.00**
JAFFE	71.57	67.62 ± 8.49 0.00	73.56 ± 4.83 0.00	64.79 ± 4.08 0.00	75.48 ± 1.63 0.00	74.98 ± 2.15 0.00	73.93 ± 2.85 0.00	75.75 ± 2.53 0.00	77.77 ± 1.87 0.00	75.46 ± 1.61 0.00	**79.29** **± 2.24** **1.00**
AR	30.26	25.29 ± 2.89 0.00	29.05 ± 1.19 0.00	**34.22** **± 2.70** **0.05**	30.87 ± 0.35 0.00	32.34 ± 1.52 0.00	31.06 ± 2.14 0.00	34.84 ± 1.90 0.04	34.19 ± 2.52 0.02	34.12 ± 1.60 0.00	**36.11** **± 0.75** **1.00**
COIL	59.17	45.60 ± 6.16 0.00	51.50 ± 5.38 0.00	50.84 ± 3.76 0.00	31.40 ± 16.89 0.00	44.22 ± 6.33 0.00	56.94 ± 3.43 0.00	59.20 ± 3.28 0.00	59.53 ± 4.01 0.03	57.96 ± 2.27 0.00	**60.93** **± 2.50** **1.00**
LUNG	72.46	58.97 ± 5.24 0.00	70.42 ± 3.41 0.00	71.58 ± 5.85 0.00	65.46 ± 3.88 0.00	75.52 ± 1.57 0.00	73.49 ± 3.43 0.00	77.35 ± 2.62 0.00	77.86 ± 3.12 0.00	77.83 ± 2.70 0.00	**81.93** **± 1.63** **1.00**
TOX	43.65	40.25 ± 0.65 0.00	43.10 ± 1.86 0.00	39.28 ± 0.49 0.00	47.14 ± 0.75 0.00	38.28 ± 1.64 0.00	39.93 ± 1.13 0.00	47.67 ± 0.83 0.00	43.96 ± 1.56 0.00	47.38 ± 1.93 0.00	**49.17** **± 0.67** **1.00**
Average	58.24	49.43	52.11	57.92	55.17	56.67	59.21	62.02	63.63	62.38	**66.03**

- The effect of adaptive structure learning

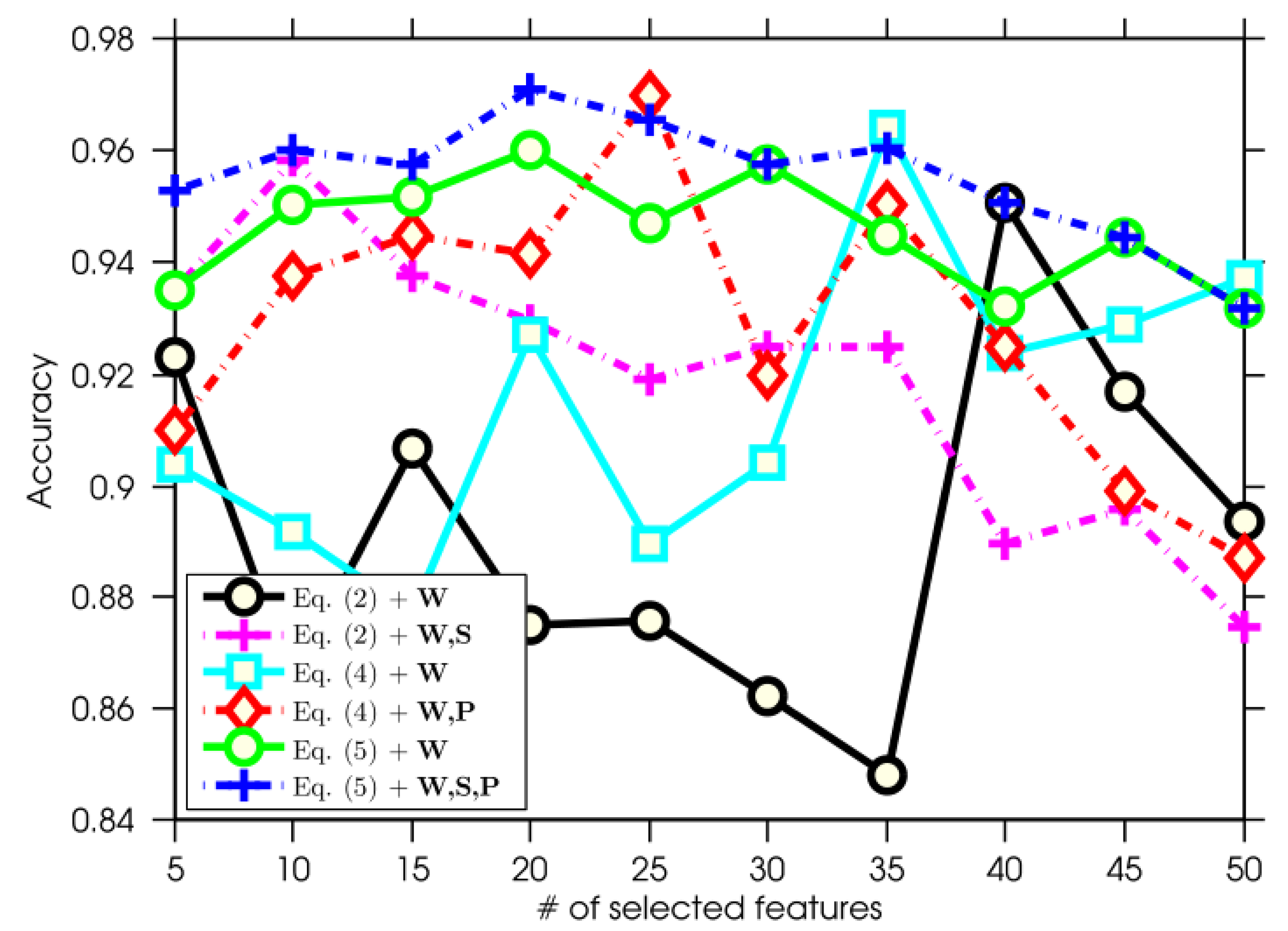

Figure 3: Clustering accuracy w.r.t. 6 different settings of FSASL on USPS200.

- Code: https://github.com/csliangdu/FSASL