Unsupervised Feature Selection via Local Total-order Preservation

The 28th International Conference on Artificial Neural Networks

Rui Ma, Yijie Wang*, Li Cheng
National University of Defense Technology, China
sherlock5204@163.com, {wangyijie, chengli09}@nudt.edu.cn

Background

- ✓ Feature selection techniques keep few informative features to alleviate the great challenges brought by high-dimensional data, such as curse of dimensionality and huge cost of computations.
- ✓ Without class label, unsupervised feature selection methods choose a subset of features that faithfully maintain the intrinsic structure of original data.
- ✓ Most methods overwhelmingly build a structure by the exact value of distance. Despite the empirical availability of high learning performance, they inevitably impose strict restrictions to the process, it causes more features to be kept for data representation.

Motivations

- ✓ Most of the existing classic feature selection methods can be interpreted from the perspective of similarity preservation. Whereas the similarity measured in high-dimensional space mightn't be qualitatively meaningful for the curse of dimensionality, especially when samples distributed in a nonlinear manifold (e.g. Fig. 1).

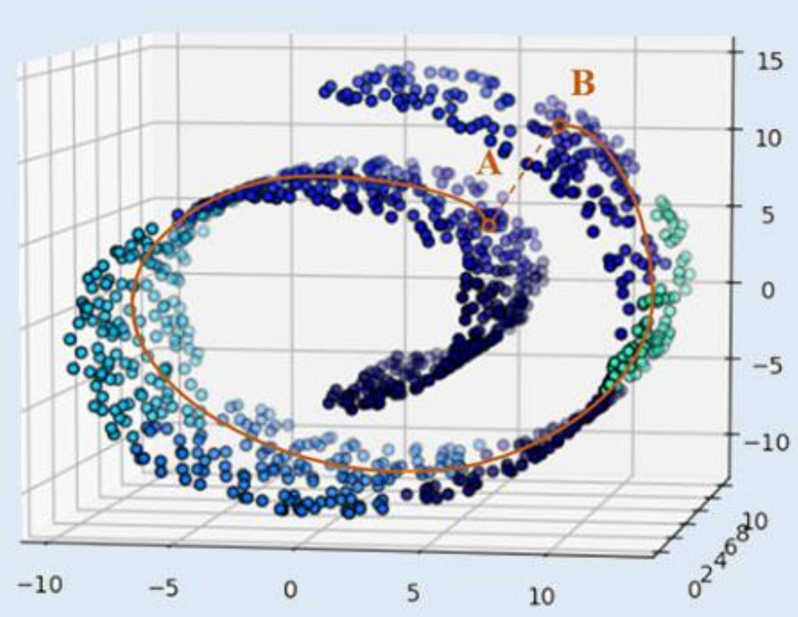

Fig. 1. An example of "Swiss roll". For points A and B on a nonlinear manifold, their Euclidean distance (length of dashed line) cannot accurately reflect their intrinsic similarity, as measured by the distance along the manifold (length of solid curve).

- ✓ In machine learning, the nearest class or cluster is referenced to label new data, so it's of vital meaning to consider the comparison of distance for learning tasks.

Basic idea

- ✓ The key challenge centers around constructing an effective geometric structure to represent the intrinsic data characteristics.
- ✓ We form total-order relation to express the comparison between instances in terms of distance. Features are selected by minimizing the differences of the relation calculated before and after feature selection.

Our method: UFSLTP

- ✓ The model of total-order relation:

[Total-order relation $x_j \geq_i x_k$]: for three different samples x_i, x_j, x_k from dataset X, there is always $d^T{}_{ij} d_{ij} \geq d^T{}_{ik} d_{ik}$, where $d_{ij} = x_i - x_j$ (e.g. Fig 2).

Before feature selection:

$$d_{ijk} = d_{ij}{}^{\mathrm{T}} d_{ij} - d_{ik}{}^{\mathrm{T}} d_{ik}$$

$$p_{ijk} = P(j \leq_i k) = 1/(1 + e^{d_{ijk}/\sigma^2})$$

After feature selection:

$$d'_{ijk} = d_{ij}{}^{\mathrm{T}} diag(w) d_{ij} - d_{ik}{}^{\mathrm{T}} diag(w) d_{ik}$$

$$q_{ijk} = P\,(j \leq_i k) = 1/(1 + e^{d'_{ijk}/\sigma^2})$$

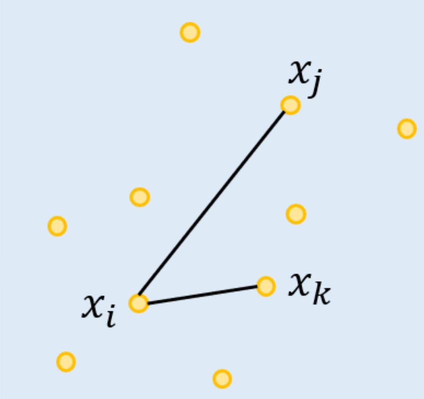

Fig. 2. Total-order relation $x_j \geq_i x_k$.

- ✓ Local Total-order relation:

Adaptive Neighbors Selection (ANS) selects k *(k=γn)* neighbors for each instance. *Uniform Neighbors Serialization (UNS)* sorts them with respect to the distance. Then, we can just calculate the total-order relation for adjacent instances *(Local Total-order)*.

- ✓ Unsupervised Feature Selection via Local Total-order Preservation (FSLTP):

$$fun(w) = \min_{w} \sum_{i=1}^{n} \sum_{j \neq i, k=j+1}^{|\mathbb{N}|-1} p_{ijk} \log \frac{p_{ijk}}{q_{ijk}} + \lambda \|w\|_1$$

$$s.t.\ 0 \leq w_t \leq 1, \forall t = 1, 2, \cdots, d$$

Experiments

- ✓ The results demonstrate that UFSLTP significantly outperforms the state-of-the-art methods. Compared to them, it averagely achieves 31.01% improvement of NMI and 14.44% in terms of Silhouette Coefficient.

Table 1. Clustering results (NMI/SC) of different algorithms over 8 datasets.

Datasets	All-Fea	LS	MCFS	UDFS	SNFS	UPFS	UFSLTP
SPECTF	0.247 0.454	0.109 0.411	0.138 0.478	0.156 0.512	0.213 0.467	0.142 0.463	**0.243** **0.513**
vehicle	0.185 0.442	0.291 0.395	0.289 0.503	**0.347** 0.445	0.299 0.418	0.319 0.362	0.346 **0.511**
segmentation	0.675 0.429	0.552 0.461	0.393 0.488	0.442 0.400	0.442 0.380	0.488 0.405	**0.556** **0.513**
optdigits	0.747 0.183	0.596 0.205	0.557 0.190	0.510 0.208	0.583 0.210	0.599 0.167	**0.649** **0.212**
Frogs	0.740 0.234	0.681 0.226	0.643 0.215	0.637 0.197	0.611 0.228	0.634 0.176	**0.684** **0.231**
colons	0.006 0.231	0.001 0.225	0.004 0.230	0.003 0.234	0.006 **0.242**	0.005 0.219	**0.008** 0.237
Yale	0.493 0.104	0.463 0.104	0.480 0.108	0.470 0.111	0.516 0.113	0.514 0.111	**0.581** **0.141**
nci9	0.426 0.050	0.428 0.052	0.437 0.052	0.436 0.044	0.428 0.052	0.438 0.042	**0.462** **0.053**
Average	0.439 0.266	0.391 0.260	0.367 0.283	0.375 0.269	0.387 0.264	0.387 0.243	**0.441** **0.301**
Average improvement by UFSLTP	10.53% 13.37%	75.00% 14.42%	28.03% 8.57%	33.61% 13.90%	17.34% 11.75%	21.52% 24.78%	*31.01%* *14.44%*