Improving Model Performance By Feature Weight Learning

颜达森

Outline

- 问题的提出
- 方法介绍
- 实验结果展示

1.问题的来源

Improving Performance of Similarity-Based Clustering by Feature Weight Learning

D.S. Yeung, *Senior Member*, *IEEE*, and
X.Z. Wang, *Member*, *IEEE*

Improving fuzzy *c*-means clustering based on feature-weight learning

Xizhao Wang [a,*], Yadong Wang [b], Lijuan Wang [a,b]

1. 目的

$$
\begin{matrix}
e_1(4.8, 5.0, 3.0, 2.0), \\
e_3(2.0, 3.0, 4.0, 5.0), \\
e_3(5.0, 5.0, 2.0, 3.0), \\
e_4(1.0, 5.0, 3.0, 1.0), \\
e_5(1.0, 4.9, 5.0, 1.0).
\end{matrix}
\quad \mathrm{x} \quad \mathrm{W} \quad = \quad
\begin{matrix}
e_1(4.8, 5.0, 3.0, 2.0), \\
e_3(2.0, 3.0, 4.0, 5.0), \\
e_3(5.0, 5.0, 2.0, 3.0), \\
e_4(1.0, 5.0, 3.0, 1.0), \\
e_5(1.0, 4.9, 5.0, 1.0).
\end{matrix}
$$

原始数据　　权重向量　　新数据

1. 相似矩阵

$e_1(4.8, 5.0, 3.0, 2.0)$,
$e_3(2.0, 3.0, 4.0, 5.0)$,
$e_3(5.0, 5.0, 2.0, 3.0)$, $*$ W $\rightarrow$ $d_{ij}^{(w)}$ $\rightarrow$ $\rho_{ij}^{(w)} = \dfrac{1}{1 + \beta * d_{ij}^{(w)}}$
$e_4(1.0, 5.0, 3.0, 1.0)$,
$e_5(1.0, 4.9, 5.0, 1.0)$.

原始数据　　距离矩阵　　相似值求法

$$\rho^{(w)} = \begin{pmatrix} 1 & 0.45 & 0.73 & 0.49 & 0.47 \\ & 1 & 0.46 & 0.45 & 0.45 \\ & & 1 & 0.46 & 0.42 \\ & & & 1 & 0.66 \\ & & & & 1 \end{pmatrix}$$

相似矩阵

1. 学习权重向量W

$$\rho \quad \begin{pmatrix} 1 & 0.45 & 0.73 & 0.49 & 0.47 \\ & 1 & 0.46 & 0.45 & 0.45 \\ & & 1 & 0.46 & 0.42 \\ & & & 1 & 0.66 \\ & & & & 1 \end{pmatrix}$$

$$E(w) = \frac{2}{n(n-1)} \times \sum_i \sum_{j \neq i} \frac{1}{2} \left(\rho_{ij}^{(w)} (1 - \rho_{ij}) + \rho_{ij} \left(1 - \rho_{ij}^{(w)} \right) \right)$$

W

$$\rho^{(w)} \quad \begin{pmatrix} 1 & 0.45 & 0.73 & 0.49 & 0.47 \\ & 1 & 0.46 & 0.45 & 0.45 \\ & & 1 & 0.46 & 0.42 \\ & & & 1 & 0.66 \\ & & & & 1 \end{pmatrix}$$

$$f(x, y) = x(1 - y) + y(1 - x) \ (1 \leq x, y \leq 1)$$

$$\frac{\partial f}{\partial x} = 1 - 2y$$

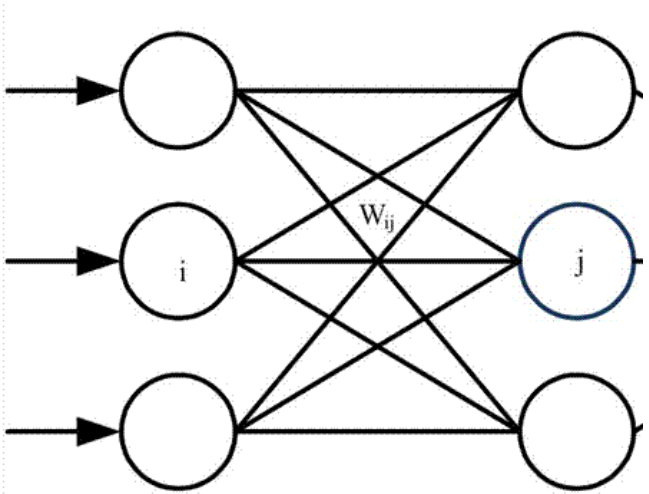
$$\frac{\partial f}{\partial x} > 0 \ \text{if} \ y < 0.5$$

$$\frac{\partial f}{\partial x} < 0 \ \text{if} \ y > 0.5$$

2. 权重矩阵

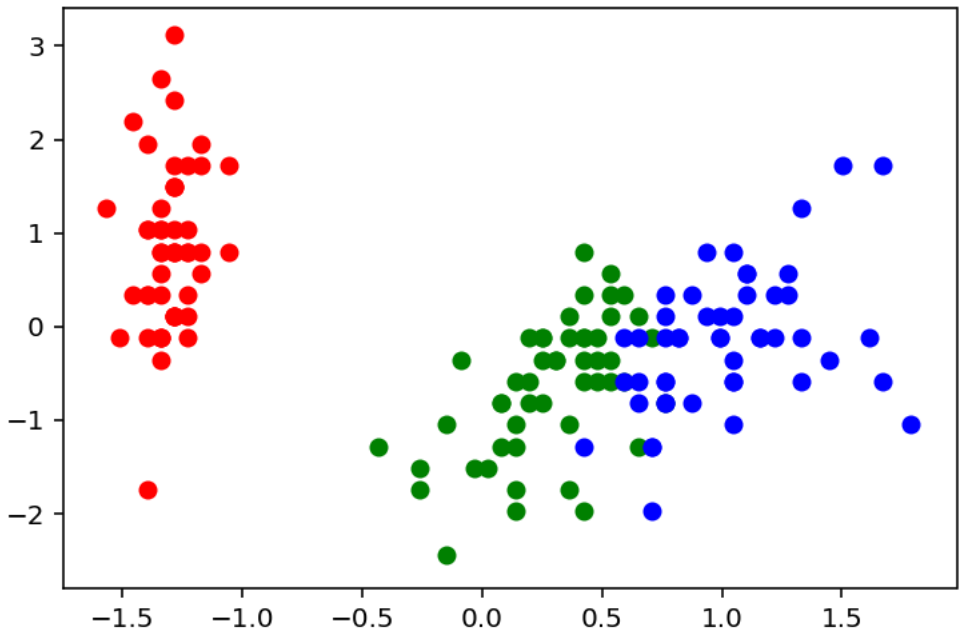
$$
\begin{matrix}
e_1(4.8, 5.0, 3.0, 2.0), \\
e_3(2.0, 3.0, 4.0, 5.0), \\
e_3(5.0, 5.0, 2.0, 3.0), \\
e_4(1.0, 5.0, 3.0, 1.0), \\
e_5(1.0, 4.9, 5.0, 1.0).
\end{matrix}
\quad \mathrm{x} \quad \mathrm{W} \quad = \quad
\begin{matrix}
e_1(4.8, 5.0, 3.0, 2.0), \\
e_3(2.0, 3.0, 4.0, 5.0), \\
e_3(5.0, 5.0, 2.0, 3.0), \\
e_4(1.0, 5.0, 3.0, 1.0), \\
e_5(1.0, 4.9, 5.0, 1.0).
\end{matrix}
$$

原始数据　　权重矩阵　　新数据

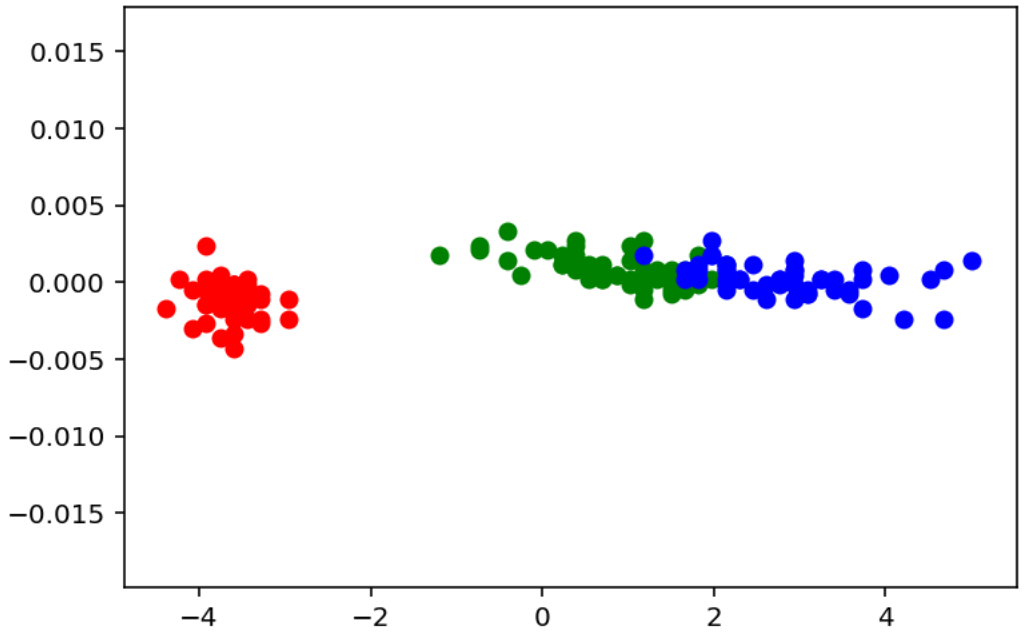


3. 实验结果

原始数据分布



新数据分布



每个数据集进行１０次实验，记录ELM在原始数据和新数据的获胜次数。

数据集	原始数据	转换后
'Autism Screening Adult Data Set .(	3	7
'Autistic Spectrum Disorder Screen	0	10
'Burst Header Packet (BHP) floodin	4	5
'Image Segmentation.csv'	3	5
'Wireless Indoor Localization Data	0	9
'abalone.csv'	8	2
'auto-mpg.csv'	8	2
'breast-cancer-D.csv'	1	9
'breast-cancer-P.csv'	4	3
'breast-cancer.csv'	0	5
'cmc.csv'	0	1
'credit.csv'	7	2
'data_banknote_authentication.csv'	6	0
'glass.csv'	4	6
'ionosphere.csv'	4	5
'messidor_features.csv'	4	6
'page-blocks.csv'	10	0
'parkinsons.csv'	4	4
'pima-indians-diabetes.data.csv'	1	6
'sonar.csv'	10	0
'waveform-+noise.csv'	1	9
'waveform.csv'	2	8
'wine.csv'	2	5
'winequality-white.csv'	10	0
'yeast.csv']	4	3
25	9	15

谢谢