DLSU RESEARCH CONGRESS 2020
"Building Resilient, Innovative, and Sustainable Societies"
June 17-19, 2020

Similarity-Preserving Feature Learning for Degraded or Incomplete Signals

Carlo Noel Ochotorena[1,*] and Yukihiko Yamashita[2]
[1] *De La Salle University, Manila, Philippines*
[2] *Tokyo Institute of Technology, Tokyo, Japan*
**Corresponding Author: carlo.ochotorena@dlsu.edu.ph*

Abstract: A large number of problems in signal processing involve signals that have been degraded by some process. In image processing, for instance, the problem of super-resolution deals with images that have been sampled with limited resolution. The objective of super-resolution is then to obtain a higher-resolution image from the original image—a severely ill-posed task. A popular approach is to estimate high-resolution patches in the image using features obtained from the input patches based on the notion that high-resolution patches that are similar to each other should also share similar features. While the conventional approach is to utilize handcrafted features extractors (e.g. gradients, Laplacians, etc.), such features may not necessarily be optimal for the given problem. In order to provide an alternative data-driven approach, this paper introduces a mathematical framework that produces a set of feature extractors from training samples comprised of the corresponding original and degraded signals, particularly in the context of super-resolution. These feature extractors are designed such that high-resolution patches with high similarity should have correspondingly high similarity in their features. The results of the training process illustrate an improvement over handcrafted features.

Key Words: feature learning; super-resolution; similarity; optimisation

1. INTRODUCTION

Signal processing drives many technological advances today. From digital assistants rooted in speech recognition and natural language processing systems to imaging systems that are embedded in everyday devices such as smartphones, signal processing has played an integral part in the improvement of such technologies. It is unsurprising, then, that the amount of research effort towards solving signal processing problems has been exponentially increasing over the past decades.

While signal processing, in itself, is a sizable field encompassing various disciplines and dealing with vastly different data types, some problems remain common across the field. One such task is that of signal reconstruction where a signal is to be estimated from incomplete or degraded measurements of the original. In most cases, such problems are generally seen as an ill-posed due to the lack of sufficient information and, as such, can only be addressed by assuming certain properties of the input signal. It is possible, for instance, to assume that some signals are slow-changing (i.e. smooth) or perhaps sparse (i.e. contains few non-zeros). The enforcement of these assumptions helps condition the problem to allow for a better estimate of the true signal.

In many domains, another approach to help reconstruct the signal is through the use of feature extraction. By assuming that extracted features are closely related to the unknown signal, it becomes possible to reconstruct the unknown signal using samples from known training signals and their corresponding features. A contextual application of this technique is example-based image super-resolution where a low-resolution image is enlarged using patches from known high-resolution images and their corresponding gradient and Laplacian features (Timofte, De, & Gool, Anchored Neighborhood Regression for Fast Example-Based Super-Resolution, 2013; Timofte, Smet, & Gool, A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution, 2014).

The use of extracted features for signal reconstruction raises certain questions:

- What feature extractors work best for a particular signal?
- How can we interpret such feature extractors?

Of the two questions, the former has been addressed through the use of hand-crafted features or machine learning techniques. By training features based on known signals and their corresponding degraded counterparts, it becomes possible to produce a good set of feature extractors. Such a task, however, is difficult to perform mathematically and is often left to heuristic learning algorithms. The use of deep learning, for instance, has been shown to be useful in arriving at feature extractors (Song, Fang, & Li, 2018). Such approaches, however, come at the cost of interpretability due to their use of a black-box approach to training.

While black-box models are generally usable in practical applications, the lack of interpretability offers little intuition on the nature of signals that can be used to empower more sophisticated techniques. In order to address this gap in interpretability, this paper re-examines feature learning from a mathematical standpoint, particularly in the context of super-resolution.

To effectively describe the proposed framework, this paper as follows: The original feature learning framework, including a similarity-preserving metric, is first described. We are specifically interested in the constraint that makes the original problem intractable. Through the use of carefully introduced auxiliary variables, we then highlight an equivalent, but tractable, framework that can be used in place of the original. We demonstrate that the proposed framework arrives at a more effective set of feature extractors based on the similarity-preserving metric.

2. FEATURE LEARNING

2.1 Example-based super-resolution

Without any goal, it is possible to arrive at an infinite number of feature extractors for a given signal. Such extractors, however, would be meaningless for the task at hand. In order to arrive at a relevant set of feature extractors, it is, therefore, necessary to quantify the relevance mathematically.

To better understand this, we, again, focus on the context of super-resolution. In the said problem, a lower-resolution image must be enlarged to form a higher-resolution image. If we utilise an upsampling factor of 2, for instance, a 1000x500 image would become a 2000x1000 image. Equivalently, if we divide the input image into patches with 3x3 pixels, we will be attempting to reconstruct the 6x6 patches through super-resolution.

Formally, we can describe low-resolution patches by collecting the pixel information into a vector $\boldsymbol{y}_i$ where the subscript is used to denote a patch located at index i. For each given patch, we can search for a corresponding estimate of the reconstructed

patch $\hat{\boldsymbol{x}}_i$ using an interpolation matrix $\boldsymbol{M}$ such that:

$$\hat{\boldsymbol{x}}_i = \boldsymbol{M}\boldsymbol{y}_i \qquad (Eq.1)$$

Our goal in super-resolution is then to find a suitable interpolation matrix that minimises the following objective:

$$\underset{\boldsymbol{M}}{argmin} \ \|\boldsymbol{x}_i - \boldsymbol{M}\boldsymbol{y}_i\|_2^2 \qquad (Eq.2)$$

where $\boldsymbol{x}_i$ describes the ground truth high-resolution patch used during training.

While we could readily describe a single interpolation matrix for all patches, such a matrix will be overly generalised and will be unable to handle differences in patch structures. It will, in fact, be roughly equivalent to naïve interpolation techniques (e.g. bilinear, bicubic, spline interpolation) used for images. Previous studies have shown that it is, in fact, more useful to describe an interpolation matrix for a *local* region of the high-dimensional patch space (Timofte, De, & Gool, Anchored Neighborhood Regression for Fast Example-Based Super-Resolution, 2013). In image patches, we can say that patches belong to the same locality if their features are similar. By dividing space into localities, we can more readily handle the properties of each of these localities (Roweis & Saul, 2000). For instance, we can construct a unique interpolator for each locality k as follows:

$$\underset{\boldsymbol{M}_k}{argmin} \ \|\boldsymbol{x}_i - \boldsymbol{M}_k\boldsymbol{y}_i\|_2^2 \quad \forall \boldsymbol{y}_i \in k \qquad (Eq.3)$$

In the specific context of super-resolution, however, researchers have found that estimating the residual relative to a naïve interpolation technique is more effective than explicitly estimating the high-resolution patch (Timofte, De, & Gool, Anchored Neighborhood Regression for Fast Example-Based Super-Resolution, 2013; Timofte, Smet, & Gool, A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution, 2014). In such a case, we can designate patches from a bicubic-interpolated image as $\tilde{\boldsymbol{x}}_i$ and describe a residual relative to the true ground truth:

$$\boldsymbol{r}_i = \boldsymbol{x}_i - \tilde{\boldsymbol{x}}_i \qquad (Eq.4)$$

such that our interpolation matrix is now tasked with finding the residuals, instead:

$$\underset{\boldsymbol{M}_k}{argmin} \ \|\boldsymbol{r}_i - \boldsymbol{M}_k\boldsymbol{y}_i\|_2^2 \quad \forall \boldsymbol{y}_i \in k \qquad (Eq.5)$$

2.2 Similarity-preserving features

In order for such local interpolators to be effective, it is vital for the localities themselves to be meaningful. Specifically, we adhere to the following criteria:

- Localities must be decided based on available information (i.e. from the input image and not the ground truth)
- Within each locality, the *ground truth* residuals must be similar to each other.

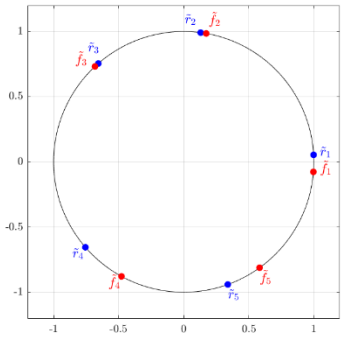
The first of these two is trivial and ensures that whatever features we use to divide the space can be extracted in a real system where no ground-truth data is available. The latter criterion, on the other hand, encourages that, even in the absence of the ground truth data, each locality should encourage a relationship among the hidden true data in that locality.

This criteria can be better visualised in Fig. 1 where we consider a hypothetical set of features and ground truth residuals. Note that since image patches naturally have varying contrast even with the same underlying structure, we first normalise the residuals to unit norm:

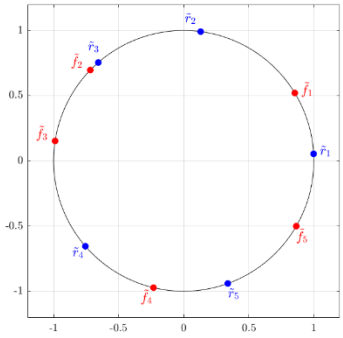
$$\tilde{\boldsymbol{r}}_i = \frac{\boldsymbol{r}_i}{\|\boldsymbol{r}_i\|_2}$$

An effect of this normalisation is that all residuals now reside on the surface of a unit hypersphere in high-dimensional space. A two-dimensional visualisation of this can be made using a unit circle with as shown in Fig. 1. We can now consider a set of hypothetical features in this representation. Much like the residuals, it is also useful to represent extracted features in normalised form $\tilde{\boldsymbol{f}}_i$.

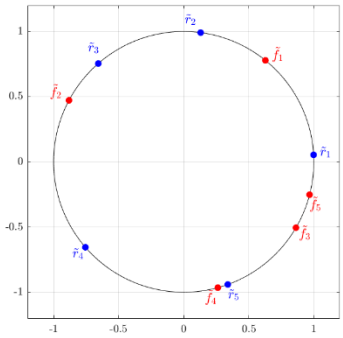
In our first example, the features are located close to the ground truth residuals and preserve the relationship between these points. This makes such features a good candidate.



(a) Good feature extraction.



(b) Good feature extraction (rotated).



(c) Bad feature extraction.

Fig. 1. Comparison of the quality of extracted features in a low-dimensional space.

Moving to the second example in Fig. 1b, we find that the features are no longer in close proximity to their corresponding residuals. While it may seem like a poor candidate for feature extraction, a closer inspection reveals a different story. Even though the distance of the ground truth data from their respective features is large, the relationship (i.e. angular distances) between any two residuals and their corresponding features is preserved. This makes this second example, likewise, a suitable candidate as the features can be used to discriminate patches even in the absence of ground truth data.

The final example in Fig. 1c, on the other hand, demonstrates a poor candidate set of features. In this last case, there is no clear preservation of the relationship between ground truth data, making the extracted features useless in localising the space.

The above illustrations highlight what we are interested in—a set of feature extractors that preserve the relationship (or similarity) of the ground truth data. Mathematically, we can define this as:

$$\underset{\boldsymbol{E},\boldsymbol{W}}{\operatorname{argmin}} \frac{1}{2}\left\|\widetilde{\boldsymbol{R}}^{\mathrm{T}}\widetilde{\boldsymbol{R}} - \boldsymbol{W}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{W}\right\|_2^2 \quad \text{s.t. } \boldsymbol{E}^{\mathrm{T}}\boldsymbol{y}_i w_i = 1 \qquad (Eq.6)$$

where $\boldsymbol{E}$ is a matrix with each column describing a unique feature extractor, $\boldsymbol{Y}$ is the collection of low-resolution training patches, and $\boldsymbol{W}$ is a diagonal matrix of weights that ensure that the unit-norm constraint is enforced.

The above objective can be interpreted as a "similarity-preserving" metric in that it is designed to preserve the inner products between residuals ($\widetilde{\boldsymbol{R}}^{\mathrm{T}}\widetilde{\boldsymbol{R}}$) and the corresponding inner products between features ($\widetilde{\boldsymbol{F}}^{\mathrm{T}}\widetilde{\boldsymbol{F}} = \boldsymbol{W}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{W}$). It should be noted that a similar metric has been used in feature *selection* techniques where a suitable *combination* of handcrafted features must be chosen to suit the given problem (Zhao, Wang, Liu, & Ye, 2013).

In this work, we introduce an additional term designed to promote sparsity in the extracted features:

$$\underset{\boldsymbol{E},\boldsymbol{W}}{\operatorname{argmin}} \frac{1}{2}\left\|\widetilde{\boldsymbol{R}}^{\mathrm{T}}\widetilde{\boldsymbol{R}} - \boldsymbol{W}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{W}\right\|_2^2 + \alpha\left\|\boldsymbol{E}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{W}\right\|_1 \quad \text{s.t. } \boldsymbol{E}^{\mathrm{T}}\boldsymbol{y}_i w_i = 1 \qquad (Eq.7)$$

2.3 Problem relaxation

While the objective function and metric described in Eq. 7 completely captures our earlier criteria, it is difficult to realise a solution for the said formulation. On one hand, you have a nonlinear optimisation problem brought about by the covariance matrix $\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}$, further tied to an ℓ_1 norm that does not have a closed-form solution. This structure, in itself, makes the problem difficult to solve. On the other hand, the presence of the normalisation weights $\boldsymbol{W}$ that must be solved simultaneously with $\boldsymbol{E}$ makes the problem even more intractable.

To reduce the complexity of the problem, we introduce auxiliary variables under an augmented Lagrangian constraint that can be iteratively optimised using the ADMM technique (Boyd, Parikh, Chu, Peleato, & Eckstein, 2011). The new objective takes on the form:

$$\begin{aligned}\underset{\boldsymbol{E},\boldsymbol{S_1},\boldsymbol{S_2},\boldsymbol{W}}{\operatorname{argmin}} \frac{1}{2}\left\|\widetilde{\boldsymbol{R}}^{\mathrm{T}}\widetilde{\boldsymbol{R}} - \boldsymbol{S_1}^{T}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{W}\right\|_2^2 + \alpha\|\boldsymbol{S_2}\|_1 \\ + \frac{\mu}{2}\left\|\boldsymbol{S_1} - \boldsymbol{E}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{W} - \boldsymbol{\Gamma_1}\right\|_2^2 \\ + \frac{\mu}{2}\left\|\boldsymbol{S_2} - \boldsymbol{E}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{W} - \boldsymbol{\Gamma_2}\right\|_2^2 \\ \text{s.t. } \boldsymbol{E}^{\mathrm{T}}\boldsymbol{y}_i w_i = 1 \qquad (Eq.8)\end{aligned}$$

where $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are the auxiliary variables and $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are their corresponding Lagrangian multipliers that have been integrated as an additive term in the norm expression. While this new objective may appear to be substantially more complicated, it allows us to break the full problem into simpler subproblems.

2.4 $\boldsymbol{S}_1$ subproblem

The first of our subproblems deals with solving for S_1 which can then be solved for each patch (i.e. column) using the method of Lagrangian multipliers:

$$\begin{bmatrix}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{W}^2\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{E} + \mu\boldsymbol{I} & \boldsymbol{E}^{\mathrm{T}}\boldsymbol{y}_i w_i \\ w_i\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{E} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{s}_{1,i} \\ \lambda_i\end{bmatrix} = \begin{bmatrix}\boldsymbol{E}^{T}\boldsymbol{Y}\boldsymbol{W}\widetilde{\boldsymbol{R}}^{T}\boldsymbol{r}_i + \mu\left(\boldsymbol{E}^{T}\boldsymbol{y}_i w_i + \boldsymbol{\gamma}_{1,i}\right) \\ 1\end{bmatrix} \quad (Eq.9)$$

This may also be carried out more efficiently on the full training set using an LDL decomposition update.

2.5 $\boldsymbol{S}_2$ subproblem

The second subproblem deals with the sparsity of the features and can be reduced to an elementwise soft-thresholding problem:

$$\boldsymbol{S}_2 = \mathcal{T}_{\alpha/\mu}(\boldsymbol{E}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{W} + \boldsymbol{\Gamma}_2) \qquad (Eq.10)$$

where the thresholding function $\mathcal{T}_t(x)$ can be defined as:

$$\mathcal{T}_t(x) = \begin{cases}\operatorname{sign}(x)(|x| - t) & |x| > t \\ 0 & |x| \le t\end{cases} \qquad (Eq.11)$$

2.6 $\boldsymbol{E}$ and $\boldsymbol{W}$ subproblem

The final subproblem is substantially more difficult due to the simultaneous optimisation of two variables but can be reduced by introducing an auxiliary variable v_i that constrained to $\boldsymbol{s}_{1,i}^{\mathrm{T}}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{y}_i$. The resulting problem can be addressed using vectorisation techniques:

$$\boldsymbol{b} = \operatorname{vec}\left(\widetilde{\boldsymbol{Y}}\widetilde{\boldsymbol{R}}^{\mathrm{T}}\widetilde{\boldsymbol{R}}\boldsymbol{S}_1^{\mathrm{T}} + \mu\widetilde{\boldsymbol{Y}}(\boldsymbol{S}_1 + \boldsymbol{S}_2 - \boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)^{\mathrm{T}} + \beta\left(\sum_i (v_i - \nu_i)\boldsymbol{y}_i\boldsymbol{s}_{1,i}^{\mathrm{T}}\right)\right) \qquad (Eq.12)$$

$$\mathrm{vec}(\boldsymbol{E}) = \left[\left(\left(\boldsymbol{S}_1\boldsymbol{S}_1^{\mathrm{T}} + 2\mu\boldsymbol{I}\right) \otimes \widetilde{\boldsymbol{Y}}\widetilde{\boldsymbol{Y}}^{\mathrm{T}}\right) + \beta\sum_i \boldsymbol{s}_{1,i}\boldsymbol{s}_{1,i}^{\mathrm{T}} \otimes \boldsymbol{y}_i\boldsymbol{y}_i^{\mathrm{T}}\right]^{-1} \boldsymbol{b} \quad (Eq.\,13)$$

with a non-linear but one-dimensional search for v_i as follows:

$$\beta v_i - \left(\boldsymbol{s}_{1,i}^{\mathrm{T}}\boldsymbol{f}_i + \nu_i\right) + \frac{1}{\beta}\left[\tilde{\boldsymbol{r}}_i^{\mathrm{T}}\left(\widetilde{\boldsymbol{R}}\boldsymbol{S}_1^{\mathrm{T}}\boldsymbol{f}_i\right) + \mu\left(\boldsymbol{s}_{1,i} + \boldsymbol{s}_{2,i} - \boldsymbol{\gamma}_{1,i} - \boldsymbol{\gamma}_{2,i}\right)^{\mathrm{T}}\boldsymbol{f}_i\right]\frac{1}{v_i^2} - \frac{1}{\beta}\left[\boldsymbol{f}_i^{\mathrm{T}}\left(\boldsymbol{S}_1\boldsymbol{S}_1^{\mathrm{T}}\boldsymbol{f}_i\right) + 2\mu\boldsymbol{f}_i^T\boldsymbol{f}_i\right]\frac{1}{v_i^3} = 0 \quad (Eq.\,14)$$

where $\boldsymbol{f}_i = \boldsymbol{E}^{\mathrm{T}}\boldsymbol{y}_i$. Note that while the above equation is non-linear, its one-dimensional nature allows for a tractable and efficient solution using Newton-based solvers.

3. EXPERIMENTAL RESULTS

Given the framework described above, it becomes possible to obtain a new set of features suitable for a specified training set. To validate the proposed framework, we collected 1 million non-smooth patches from the DIV2K database (Agustsson & Timofte, 2017) and iteratively applied our proposed framework to the said patches. An initial set of gradient and Laplacian features were used to train the new features and the subsequent unconstrained metric from Eq. 6 was used to measure the performance of the new features. To better interpret the results, the metric was expressed relative to that obtained from the original gradient and Laplacian features and used to score the new set of features. The results over the training iterations can be seen in Fig. 2. Note that a lower relative score is better.

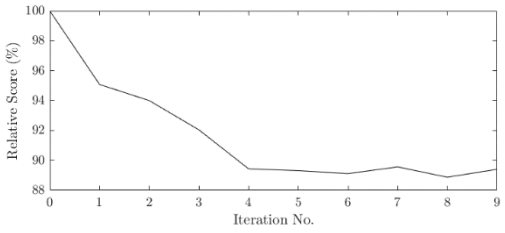


Fig. 2. Relative scores of each successive set of features. A lower score indicates a better preservation of the relationship between residuals.

The results obtained during the training process clearly demonstrate an improvement over the initial feature set. An interesting observation from these results is that the gradient and Laplacian features of the image are already a good candidate set of features as demonstrated in previous studies. Nonetheless, our proposed framework highlights how further improvements can be obtained by carefully refining such features.

4. CONCLUSIONS

Interpretable feature-learning is a highly complex and non-linear task that impacts various fields of signal processing. This work presents a mathematical framework that reduces the complexity in order to tractably solve the problem. Our experimental results show that the proposed framework is effective in handling the feature learning problem in a reasonable amount of time (several hours for a training size of 1 million patches). Beyond this study, it is still necessary to analyse the convergence of the proposed technique and evaluate the effects of additional constraints.

5. ACKNOWLEDGEMENTS

This work was supported by the De La Salle University (DLSU) University Research Coordination Office (URCO) under Project No. 59-F-U-3TAY18-3TAY19.

6. REFERENCES

Agustsson, E., & Timofte, R. (2017). NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (pp. 1122-1131).

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers.

Jiang, J., Ma, X., Chen, Lu, T., Wang, Z., & Ma, J. (2017). Single Image Super-Resolution via Locally Regularized Anchored Neighborhood Regression and Nonlocal Means. IEEE Transactions on Multimedia, 19(1), 15-26.

Kiku, D., Monno, Y., Tanaka, M., & Okutomi, M. (2016). Beyond Color Difference: Residual Interpolation for Color Image Demosaicking. IEEE Transactions on Image Processing, 25(3), 1288-1300.

Li, Z., & Tang, J. (2015). Unsupervised Feature Selection via Nonnegative Spectral Analysis and Redundancy Control. IEEE Transactions on Image Processing, 24(12), 5343-5355.

Roweis, S., & Saul, L. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science, 290(5500), 2323-2326.

Song, W., Fang, L., & Li, S. (2018). Similarity-Preserving Deep Features for Hyperspectral Image Classification. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, (pp. 3595-3598).

Timofte, R., De, V., & Gool, L. (2013). Anchored Neighborhood Regression for Fast Example-Based Super-Resolution. 2013 IEEE International Conference on Computer Vision, (pp. 1920-1927).

Timofte, R., Smet, V., & Gool, L. (2014). A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution. asian conference on computer vision, 9006, 111-126.

Zhao, Z., Wang, L., Liu, H., & Ye, J. (2013). On Similarity Preserving Feature Selection. IEEE Transactions on Knowledge and Data Engineering, 25(3), 619-632.