TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Feature Dimensionality Reduction with Graph Embedding and Generalized Hamming Distance

Honglei Zhang, Moncef Gabbouj
Signal Processing Tampere University of Technology, Finland

1 Introduction

- Feature dimensionality reduction using graph embedding paradigm
- Intrinsic and penalty graph [2]
- Graph embedding unifies PCA, LDA, Isomap and many other methods
- For multilabel problems, how about correlation between the labels?

2 Previous Work

- Let $L = \{L_1, L_2, \cdots, L_q\}$ be the set of labels. Let $X = \{x_1, x_2, \cdots, x_N\}$ be the set of the samples where $x_i \in R^M$. Let $Y = \{y_1, y_2, \cdots, y_N\}$ be the labels, where $y_i \in \{0,1\}^q$.
- Our target is to learn a linear projection $z = Wx$ where $W \in R^{P\times M}$, $P < M$.
- Objective function

$$J = \sum_{i,j,i\neq j} \|Wx_i - Wx_j\|^2 A_{ij}. \qquad (1)$$

- The regularization term

$$x^T W^T BWx = I. \qquad (2)$$

- For PCA, $A_{ij} = \frac{1}{N}$ and $B = I$; For LDA, $A_{ij} = \delta(y_i, y_j)$ and $B = 1 - \frac{1}{N}ee^T$.
- When $B = I$, the solution of this optimization problem can be obtained by solving the following eigenvalue problem

$$\tilde{L}w = \lambda w,$$

where $\tilde{L} = X^T LX$ and L is the Laplacian matrix of the intrinsic graph [1]. By keeping the first P eigenvectors of matrix $\tilde{L}$ with the largest eigenvalues, we get the matrix W^*.
- For the multilabel problems, data points sharing many common labels should be close to each and data points that do not share common labels should be separated far away
- Euclidean distance: $A_{ij} = \|y_i - y_j\|^2$
- Hamming distance:

$$A_{ij} = count(y_i \oplus y_j), \qquad (3)$$

where $\oplus$ is the XOR operator and $count(\cdot)$ calculates the number of 1s. Note, hamming distance calculate number of labels that differs in y_i and y_j

3 Methodology

3.1 Normalized mutual information of labels

For two random variables X and Y, mutual information is defined as

$$I(X;Y) = \sum_y \sum_x p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \qquad (4)$$

Normalized mutual information is defined as

$$NI(X,Y) = \frac{I(X;Y)}{\min(H(X), H(Y))}, \qquad (5)$$

where $H(X)$ and $H(Y)$ are the marginal entropies of variable X and Y. Given the multilabel data and take each label L_i as a random variable, we have

$$p(L_i) = \frac{1}{N} \sum_{k=1}^{N} y_k(i), \qquad (6)$$

$$p(L_i, L_j) = \frac{1}{N} \sum_{k=1}^{N} y_k(i) y(j). \qquad (7)$$

3.2 Generalized hamming distance

Hamming distance defined in 3 can be written as

$$A_{ij} = count(y_i \vee y_j) - \langle y_i, y_j \rangle, \qquad (8)$$

where "$\vee$" is the "or" operator of two binary vectors. The inner product of two vectors y_i and y_j with nonorthogonal basis is defined as

$$\langle y_i, y_j \rangle = \sum_l \sum_m y_i(l) y_j(m) \langle \mathbf{e}_l, \mathbf{e}_m \rangle, \qquad (9)$$

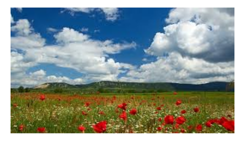
where $\mathbf{e}_l$ and $\mathbf{e}_m$ are the basis vectors. From Eqs. 8 and 9, we define the generalized Hamming distance of the sample x_i and x_j with label y_i and y_j as following:

$$A_{ij} = count(y_i \vee y_j) - y_i^T F y_j, \qquad (10)$$

where F is the normalized mutual information matrix.

Theorem 1 *Generalized Hamming distance becomes Hamming distance if labels are mutually independent.*

4 Multilabel Example



flower, garden, sky, cloud, calm

child, dog, labrador, lovely

whale, ocean, fish

5 Experiments

Raking loss values of dimensionality deduction methods and ML-kNN on Yeast data

Dimension	1	2	4	8	16
PCA	**0.209**	**0.202**	0.196	0.18	**0.172**
Hamming	0.212	0.205	0.198	**0.177**	0.174
Euclid. Y	0.21	0.208	0.198	**0.177**	0.173
Euclid. X	**0.209**	0.204	0.197	0.179	0.173
GH	**0.209**	**0.202**	**0.195**	**0.177**	**0.172**

Ranking loss on NUS-WIDE 128 dataset - feature dimension is 4

Dimension	MLkNN	IBLR_ML	BRkNN	DMLkNN	RLkNN
PCA	0.098	0.106	0.128	0.097	0.1422
Hamming	**0.096**	**0.103**	**0.126**	**0.095**	0.1396
Euclid. Y	0.097	0.104	0.126	**0.095**	0.1403
Euclid. X	0.098	0.106	0.128	0.097	0.142
GH	**0.096**	**0.103**	**0.126**	**0.095**	**0.1394**

6 Conclusion

- For multilabel problems, labels are correlated
- Generalized hamming distance captures the correlation between the labels
- The results show that the proposed method consistently outperforms other dimensionality reduction methods.

References

[1] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[2] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and S. Lin. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.