Multi-label Feature Selection via Global Relevance and Redundancy Optimization

Jia Zhang[1], Yidong Lin[2], Min Jiang[1], Shaozi Li[1], Yong Tang[3], Kay Chen Tan[4]

1. Department of Artificial Intelligence, Xiamen University, China
2. School of Mathematical Sciences, Xiamen University, China
3. School of Computer Science, South China Normal University, China
4. Department of Computer Science, City University of Hong Kong, Hong Kong

Abstract

Information theoretical based methods have attracted a great attention in recent years, and gained promising results to deal with multi-label data with high dimensionality. However, most of the existing methods are either directly transformed from heuristic single-label feature selection methods or inefficient in exploiting labeling information. Thus, they may not be able to get an optimal feature selection result shared by multiple labels. In this paper, we propose a general global optimization framework, in which feature relevance, label relevance (i.e., label correlation), and feature redundancy are taken into account, thus facilitating multi-label feature selection. Moreover, the proposed method has an excellent mechanism for utilizing inherent properties of multi-label learning. Specially, we provide a formulation to extend the proposed method with label-specific features. Empirical studies on twenty multi-label data sets reveal the effectiveness and efficiency of the proposed method.

Approach

we propose a new multi-label feature selection method via global relevance and redundancy optimization, named GRRO. In addition, we also give an extension of GRRO to conduct feature selection with label-specific features, named GRRO-LS.

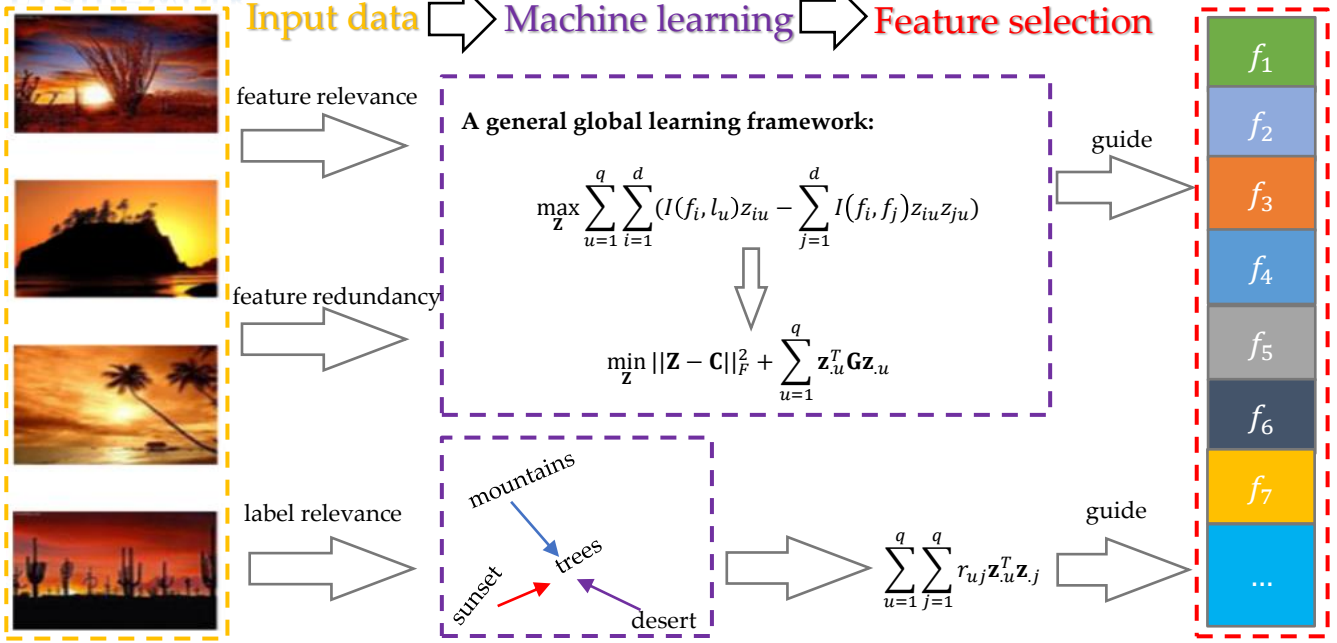
1. Multi-label feature selection

Considering that many feature evaluation criteria are proposed to maximize feature relevance and minimize feature redundancy, we take the advantage to measure feature importance. Different to the heuristic search, we propose to achieve the purpose with optimization. Furthermore, we learn the global label relevance by exploiting second-order label correlation.

2. Label-specific feature selection

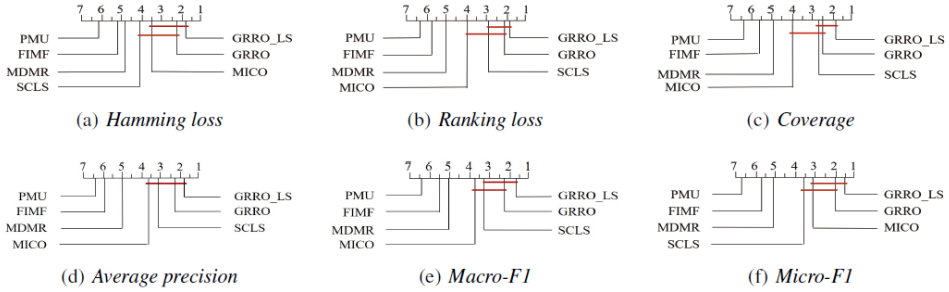
The global optimization result, generated by GRRO, is utilized as the *priori* knowledge. We analyze discriminative features for each label, and these features are specified as the label-specific features with respect to the corresponding label. Thus, we update the weight matrix with the weight information of label-specific features, thus achieving the purpose.

Framework



Experiment

Performance Evaluation: The proposed methods. i.e., GRRO-LS and GRRO, rank 1st and 2nd respectively, and GRRO-LS can achieve highly competitive performance against the selected comparing methods.



Comparison of the control method against comparing methods with the Nemenyi test (CD = 2.0146 at 0.05 significance level)

Efficiency Evaluation:

1. GRRO performs the best in terms of average ranking (Ave. Rank.). Thus, GRRO has the overall best performance in time cost.
2. Theoretically, the proposed methods GRRO-LS and GRRO have the similar result on running time.

Running Time Comparison

Data set	PMU	MDMR	FIMF	SCLS	MICO	GRRO
Bibtex	–	–	–	**194.30**	–	200.61
Birds	40.02	37.65	**0.66**	1.51	1.39	0.89
Corel5k	–	–	–	95.22	42.18	**22.95**
Corel16k001	–	–	743.89	87.20	50.02	**32.46**
Corel16k002	–	–	937.84	92.20	50.04	**33.23**
Emotions	2.83	2.73	0.06	0.32	0.64	**0.05**
Genbase	295.57	263.61	5.95	**3.04**	11.01	10.83
Image	23.26	22.39	**0.09**	2.47	2.80	1.16
Langlog	–	–	49.99	16.65	100.89	**14.15**
Medical	876.15	713.29	21.50	**13.12**	193.50	20.90
Slashdot	808.58	717.25	**8.47**	22.48	56.61	36.07
Yeast	30.21	27.21	0.25	1.22	0.95	**0.23**
Arts	345.59	304.52	**4.05**	8.22	11.39	5.42
Business	387.55	334.92	5.19	8.12	9.72	**4.98**
Entertainment	383.35	344.17	**3.78**	10.81	28.30	10.01
Health	589.30	502.54	**8.23**	12.01	24.68	9.50
Recreation	382.83	339.72	**3.92**	10.23	23.67	9.12
Reference	796.20	671.31	**11.51**	16.01	56.24	15.98
Science	944.48	765.26	15.54	16.03	46.30	**14.28**
Social	–	–	**20.82**	22.70	148.17	27.57
Ave. Rank.	6.00	5.00	2.05	2.65	3.55	**1.75**

Conclusion

We developed information theoretical based methods for multi-label feature selection. Our main contribution is to propose a general global optimization framework incorporating feature relevance, feature redundancy, and label correlation. In addition to label correlation exploitation, the proposed method is capable for exploiting the other properties of multi-label learning to further improve the performance, such as label-specific features. Experiments on twenty benchmark data sets in terms of six evaluation metrics showed that the proposed method can significantly improve the performance with feature selection.