ReliefF-based Multi-label Feature Selection

Yaping Cai[1], Ming Yang[2], Yang Gao [3] and Hujun Yin[4]

[1] [2]*School of Computer   Science and Technology,*
Nanjing Normal University, Nanjing210023. PR.China
[3] *State Key Laboratory for Novel Software Technology, Nanjing University,*
Nanjing210023. PR.China
[4] *School of Electrical and Electronic Engineering,*
The University of Manchester, Manchester, M13 9PL,UK

Abstract

In recent years, multi-label learning has been used to deal with data attributed to multiple labels simultaneously and has been increasingly applied to various applications. As many other machine learning tasks, multi-label learning also suffers from the curse of dimensionality; so extracting good features using multiple labels of the datasets becomes an important step prior to classification. In this paper, we study the problem of multi-label feature selection for classification and have proposed a method based on single label feature selection ReliefF, termed ML-ReliefF, to select discriminant features in order to boost multi-label classification accuracy. Compared to other multi-label feature selection methods that only consider the relationship between pairwise classes, the proposed method introduces the concept of label set to further consider the relationship among more than two labels, modifies the regulation of the nearest neighbors computation reflecting the influence between samples and multiple labels, and considers and adds the similarity between samples to reinforce the effect. With the classifier, ML-kNN, experiments on five different datasets show that the proposed method is effective in removing irrelevant or redundant features and the selected features are more discriminant for classification.

Keywords: Multi-label learning, Feature selection, Multi-label classification.

1. Introduction

In traditional task of single-label learning, each instance is only associated with a single class label. However, in many real-world problems, one object can possess multiple concepts simultaneously, so leading to the so called multi-label learning[1-16]. The multi-label learning has soon received increased attention and is now applicable to a wide variety of domains, including text classification[3-5], gene function classification[6-8], image annotation[9], video automatic annotation[10, 11], web mining[12, 13] and so on. In the last decade, a number of multi-label classification algorithms[3-17] have been developed for multi-label data, such as Binary Relevance method[9], lazy method[15], LEAD method[17].

The curse of dimensionality[18] always causes serious problems when learning with high-dimensional data. PCA (principal component analysis)[19] and LDA (linear discriminant analysis)[20] are two classic dimensionality reduction methods. To extend to the multi-label learning, the MLSI method[21], the MDDM method[22] and some other methods[23, 24] have been proposed. However, these dimensionality reduction methods transform the high-dimensional feature space into a new, lower dimensional space with a linear or non-linear projection, which may make the features lose their original semantic information so making it difficult to remove irrelevant or redundant features. Feature

selection is a process of selecting a subset of well distinguished features from all the original features for classification. Feature selection methods can be broadly divided into two categories: filter method[18, 25] and wrapper method. In a multi-label case, Zhang incorporated feature selection mechanisms into MLNB to improve the classification performance[26]. Kong, et al. improved the classical ReliefF and F-statistic algorithms and adapted the single-label methods to multi-label data[27]. Though there exist several feature selection methods for multi-label data, multi-label feature selection is still a challenge problem. The reason is that in the multi-label case, one data point may be attributed to multiple labels simultaneously, features are not at ease to decide whether they are useful for all the multiple concepts[27].

ReliefF[25] is a classical feature selection algorithm. It utilizes the correlation between the characteristics to make similar samples close and keep heterogeneous samples apart in order to achieve the purpose of the feature selection. An extended ReliefF method, MReliefF[27], has been proposed for multi-label feature selection. It decomposes the problem into a collection of two-class problems and discussion for the cases that the "hit" and "miss" sets contain both classes is provided.

In this paper, we present a new multi-label feature selection algorithm based on ReliefF, we termed it ML-ReliefF. Some characteristics of the proposed method are highlighted below:

To study the correlation between multiple labels, we set all the labels one instance owns as a labelset of the instance. To study the correlation between the samples and the labels, we modify the calculating regulation of k nearest neighbors by using the number of same label they owns with the sample instead of the distance. To study the correlation between the samples, we add the similarities between the samples into the updating formula of feature weights.

The rest of the paper is organized as follows. In the next section, some related work is briefly introduced. The proposed ML-ReliefF method is then described in section 3 and experimental results are presented in section 4. Finally, section 5 concludes and sets out future work.

2. Related Works

2.1. Single Label ReliefF Feature Selection

Assume that there are n instances and L labels. Let $P \in R^f$ be the full set of features, $p \in P$ be a feature, $X = [x_1, x_2, ..., x_n] \in R^{n \times f}$ denote instances and let $Y = [y_1, y_2, ..., y_n] \in R^{n \times L}$ denote the instances with labels. One instance represented by $x_i \in R^f$ can be expressed as $x_i = [p_i^1, p_i^2, ..., p_i^f]$. It is associated with a set of labels by a binary vector $y_i = \{0,1\}^l$, and $y_i(l) = 1$ if x_i belongs to the l th class and $y_i(l) = 0$ otherwise. Since an instance owns multiple labels, $\sum y_i(l) \geq 1$.

For the classical ReliefF[25], the algorithm samples m instances randomly from the dataset. For each sample point $x_t (1 \leq t \leq n)$, it finds K nearest neighbors that belongs to the same class C as x_t named as Hit and for other $(L-1)$ classes (other than C), it also finds K nearest neighbors denoted as Miss(C); So the formula for updating every feature is computed as,

$$W_p = W_p - \sum_{j=1}^{K} \frac{d(p, x_t, H_j)}{m \cdot K} + \sum_{C \neq C(x_t)} \sum_{j=1}^{K} \frac{P(C)}{1 - P(C(x_t))} \cdot \frac{d(p, x_t, M_j)}{m \cdot K} \qquad (1)$$

Where W_p denotes the value of feature p, $P(C)$ is the priori probability of the label class C, and $d(p, x_t, x_j)$ is the distance between x_t and x_j on feature p (usually the Euclidian distance).

2.2. Multi-label Difficulties

As we know, in single-label feature selection, an instance has only one label. For a sample point x_t , the algorithm can straightly distinguish the other samples that have the same label with x_t, or belong to a different label C. Thus it is easy to update the feature weights by the K nearest neighbors from Hit and Miss(C), and select the top D features. However, in the multi-label problem, the label that sample point x_t is associated with is not one label class but a set of labels, and other samples that own a subset of the label set of x_t may also own some labels out of the label set.

Table 1. An Example of Multi-label Difficulties x_2 , x_3 **and** x_4 **in Table 1**

Table 2. N_{label} **of**

Label\Sample	L_1	L_2	L_3	L_4
x_1	1	0	1	0
x_2	1	1	0	1
x_3	1	0	1	1
x_4	0	1	0	1

Sample	N_{label}
x_2	1
x_3	2
x_4	0

Table 1 shows a case that x_1 belongs to L_1 and L_3 simultaneously, x_2 has the same label L_1 as x_1, but also has label L_2 and L_4, which x_1 isn't associated with, so does x_3 . Thus, how to delimit the members for Hit and Miss(C) is a challenge.

In MReliefF[27], the difficulty is overcome by decomposing the problem into a set of pairwise two-class single-label classification problems. It works well on some image annotation tasks. Now, we want to further consider about the influence that the whole label set on each instance and each other labels out of the set in terms of feature weights. Another objective is to make the method more generalized not only for image annotation tasks, but also on other multi-label tasks.

3. Multi-label ReliefF Feature Selection

3.1. Aspects of ML-ReliefF

As one instance can have more than one labels, we define the set of all related labels as the label set, and we define the Hit and Miss(C) by the number of the labels that each other sample point associated with and is the same as in the label set. Corresponding to that, we define and adapt the prior probabilities of the label set. In addition, we add the similarity between the sample points into the updating formula of feature weights to consider the impact from the nearest neighbors on each feature. Next we will introduce our method from three aspects in detail.

Definition of Hit and Miss(C): For multi-label problems, we delimit all the labels that a sample point x_t is attached with compose to a label set LS_t . For example, in Table 1 x_1 belongs to L_1 and L_3 , we site $\{l_1, l_3\}$ to be the label set LS_1 that x_1 is attached with. Then we consider the computation of the Hit and Miss(C). As the label set LS_t of x_t is definitive, we compute N_{label} for each of the other samples. N_{label} is the number of labels one remaining sample owns that appear in LS_t. Considering the case in Table 1, we choose x_1 to be the sample point and show N_{label} of the other three samples in Table2. For the Hit, the remaining samples are ordered by their N_{label} in descending, otherwise for Miss(C) he samples are ordered by the N_{label} in ascending.

Table 3. Results (a) Hit and (b) Miss(C)

Miss(L_2)		Miss(L_4)	
Sample	N_{label}	Sample	N_{label}
x_4	0	x_4	0
x_2	1	x_2	1
-	-	x_3	2

(a) Hit   (b) Miss(C)

Table 3 shows the result of Hit and Miss(C) for Table 2. We only select the top K to be the final members of the Hit and Miss(C).In summary, for the example of Table 1, assume that $K = 1$, we select the first sample of Hit and Miss(C) in Table 3, so finally the member in Hit of sample point x_1 is x_3, the member in Miss(L_2) is x_4 and in Miss(L_4) is x_4 too.

Prior probability of the label set: After we delimit the concept of label set for instance x_t, the prior probability $P(C(x_t))$ in Eqn.(1) should be reconsidered. In single label ReliefF feature selection, the $P(C(x_t))$ is computed simply as,

$$P(C(x_t)) = \frac{N_{C(x_t)}}{n} \qquad (2)$$

$N_{C(x_t)}$ is the number of the instances that have the same label with x_t, and n is the number of all instances. Because one instance is associated with a set of labels, and for other instances, one may be associated with multiple labels in the label set. So we redefine the $N_{C(x_t)}$ as N_{LS_t}, given by

$$N_{LS_t} = \bigcup_{C(x_t) \in LS_t} N_{C(x_t)} \qquad (3)$$

With N_{LS_t} in Eqn.(3) substituted into Eqn.(2), we obtain the new computation of the prior probability of the label set LS_t as,

$$P(LS_t) = \frac{N_{LS_t}}{n} = \frac{\bigcup_{C(x_t) \in LS_t} N_{C(x_t)}}{n} \qquad (4)$$

Consideration of similarity: The aim of the single label ReliefF feature selection is to select a subset of the most discriminant features by making similar samples closer and keeping heterogeneous samples further away. In multi-label ReliefF method, we want to take the similarity between the sample point x_t and other samples into consideration. That means in the updating formula of feature weights, we also compute and consider the distance as itself multiplied by the similarity between x_t and each of the corresponding K nearest neighbors in Hit and Miss(C) as,

Sample	N_{label}
x_3	2
x_2	1
x_4	0

$$\sum_{j=1}^{K} sim_{t,j} \cdot \frac{d(p, x_t, H_j)}{\sum_{j=1}^{K} sim_{t,j}} \qquad (5)$$

$$\sum_{j=1}^{K} sim_{t,j} \cdot \frac{d(p, x_t, M_j)}{\sum_{j=1}^{K} sim_{t,j}} \qquad (6)$$

$sim_{t,j}$ in Eqn.(5) gives the similarity between x_t and H_j (a member of Hit), and Eqn.(6) gives the similarity between x_t and M_j (a member of Miss(C)). Whether in the Hit and Miss(C), the neighbors that have a greater similarity with x_t are closer to x_t. We use cosine to measure the similarity, so $sim_{t,j}$ is computed as,

$$sim_{t,j} = \frac{1}{1+\cos_{t,j}} \quad (7) \qquad \cos_{t,j} = \frac{\sum_{p=1}^{f} x_t^p \cdot H_j^p}{\sqrt{\sum_{p=1}^{f}\left(x_t^p\right)^2}\sqrt{\sum_{p=1}^{f}\left(H_j^p\right)^2}} \quad (8)$$

Eqn.(8) shows the computation of $\cos_{t,j}$ between x_t and H_j, the calculation of the cosince between x_t and M_j is similar. In this way, the label information has been considered by adding correlations between the instances in multi-label feature selection.

3.2. ML-ReliefF Algorithm

From the above aspects, we come to the updating formula for the feature weights as,

$$W_p = W_p - \sum_{j=1}^{K} sim_{i,j} \cdot \frac{d\left(p, x_t, H_j\right)}{m\sum_{j=1}^{K} sim_{i,j}} + \sum_{C \neq C(x_t)} \sum_{j=1}^{K} \frac{P(C)}{1-P(C(x_t))} \cdot sim_{i,j} \cdot \frac{d\left(p, x_t, M_j\right)}{m\sum_{j=1}^{K} sim_{i,j}} \quad (9)$$

The Algorithm below shows the procedure.

Algorithm ML-ReliefF

Input:
   Instance-Feature matrix X. Instance-Label matrix Y.
   Number of samples m. Number of nearest neighbors K.
Output:
   Vector of feature weights W.
Method:
   (1) Initialize W that $W_p = 0 \left(p = 1 : f\right)$
   (2) For $t = 1 : m$
   (3)   Randomly select a sample x_t and get the corresponding label set LS_t
   (4)   Find the Hit of x_t with label set LS_t
   (5)   For $C \notin LS_t$
   (6)   Find the Miss(C) of x_t
   (7)   For $p = 1 : f$
   (8)   Update the feature weights as Eqn. (9)
   (9)   End for
   (10)   End for
   (11) End for

4. Experiment

A series of experiments were conducted on four different kinds of multi-label datasets. Table 4 summarizes the descriptions of these datasets.

Table 4. Multi-Label Dataset Description

Dataset	Type	Train	Test	Label	Dimension
Image	images	1200	800	5	294
Human	biology	1862	1244	14	440
Plant	biology	588	390	12	440
Medical	text	645	333	45	1449

Different from the conventional classification problem, the multi-label classification requires more complex evaluation metrics. We used the F1 score [27] and five evaluation metrics described in [15]. We used the multi-label classifier: ML-kNN [15] to evaluate the performance of

different feature selection algorithms, in which we set the number of the nearest neighbors in ML-kNN to 10. The reasons for choosing the ML-kNN classifier are as follows: It is a classifier with simple approach but good performance. And it is a multi-label classifier that is widely applied and used.

In the experiments, the proposed algorithm was compared with three other methods:

(1) ORI[18]: using all features (no selection); (2) single-label ReliefF[19]; (3) MDDM[22]: It is a feature extraction method. For the multi-label dimensionality reduction, there are many feature extraction methods that give good performance but few feature selection methods. We compare our method with MDDM to verify its effectiveness; (4) MReliefF: It is the first and perhaps the only existing multi-label feature selection method based on ReliefF[27]. The authors have verified the superiority of MReliefF over the BR-ReliefF and LP-ReliefF, which are based on problem transformation.

4.1. Comparison of Feature Selection Methods

In the experiments, for the ReliefF, MReliefF and ML-ReliefF, the parameter K was set to 1 on the Medical and 3 on the other three datasets. We also set parameter m (the number of samples) to 300. The latter part of this section will illustrate the reasons for the selection of these parameters.

Fig.1 shows the F1 scores of different methods with classifier ML-KNN for every 10 percentage. From the results, one can see that on the Image dataset the performances of the proposed method and other methods are almost the same; but on the other three datasets, our multi-label feature selection method performed better than the other methods on the F1 score, indicating that our multi-label feature selection method can be better suited for more types of datasets than other methods.
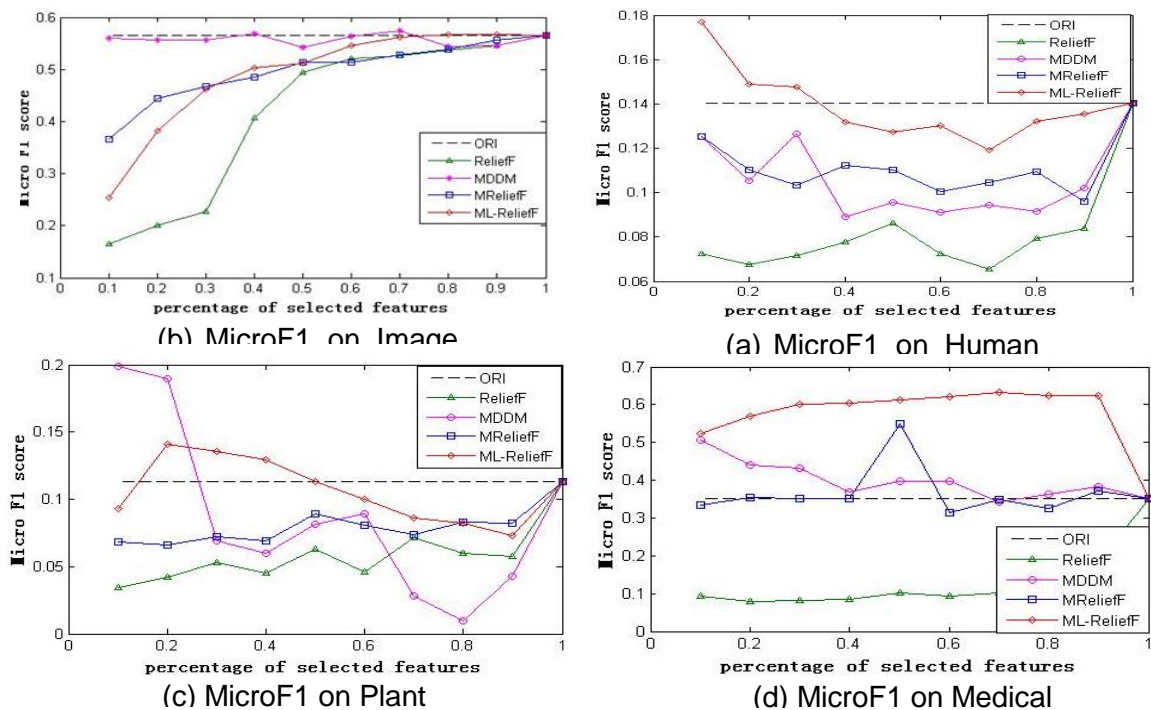

(b) MicroF1 on Image
(a) MicroF1 on Human
(c) MicroF1 on Plant
(d) MicroF1 on Medical

Figure 1. **Comparisons of Classification Results using different methods with MicroF1**

With the result of F1 score on different dimensions, we chose the part that both the MReliefF and ML-ReliefF perform well to compare their average precision. For the Image set, we chose the dimensions between 215 and 265, and for the Human, Plant and

Medical, sets we chose the   dimensions between 20 and 60, 65 and 115, 850 and 895. With each dataset, we compared the two methods every five dimensions. In Fig.2, the average precisions and comparisons between MReliefF and ML-ReliefF on four datasets are given.
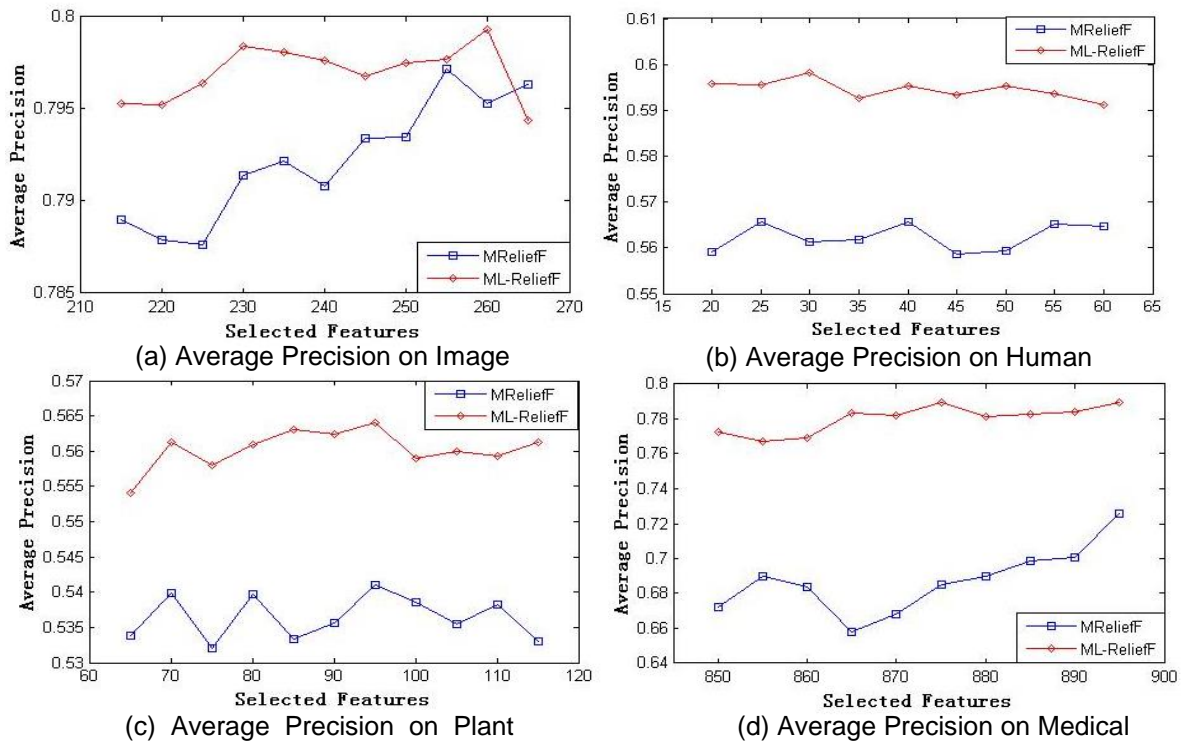

(a) Average Precision on Image

(b) Average Precision on Human

(c) Average Precision on Plant

(d) Average Precision on Medical

Figure 2. Comparisons between MRelifF and ML-ReliefF with Average Precision

From the results, one can see that on the Image dataset, the two methods have the same good performance and on the other datasets, the ML-ReliefF outperforms the MReliefF. As on the Plant and Human, the proposed method achieved 3% higher average precision than MReliefF, and about 10% on Medical.

In Table 5, we chose the selected dimensions which we regarded both methods gave good performance, as it has been shown in Fig.2. It can be seen that for the Image dataset, the proposed method works as well as the MDDM and MReliefF. As the number of dimensions increases, our multi-label feature selection method yields a much better performance, especially on the Medical dataset than the MReliefF. We attribute this to two reasons:
(1)In certain datasets, one can find that there are some labels which are only attached with one sample, so if one uses the MReliefF, one may lose information on the nearest neighbors. The proposed method, which uses labelsets, can avoid this problem to a certain degree. (2)As the number of features increases, distance can't be a good measure for computation of nearest neighbors. So we use the number of the   labels which the test sample and train samples both contains to choose the nearest neighbors which makes the similarity measure performance better.

Table 5. Compare the ML-Relief to other Methods with Different Evaluation Metrics

Dataset	Dimension	Metrics	ORI	ReliefF	MDDM	MReliefF	ML-ReliefF
Image	**260**	Hamming Loss	0.1722	0.1823	**0.1700**	0.1796	0.1712
		One-error	0.3237	0.3375	**0.3075**	0.3225	0.3100
		Coverage	0.9663	1.0215	**0.9425**	0.9625	0.9501
		Ranking Loss	0.1755	0.1813	**0.1679**	0.1712	0.1724
		Average Precision	0.7916	0.7806	**0.8008**	0.7953	0.7993
Human	**25**	Hamming Loss	0.0831	0.0841	0.0818	0.0833	**0.0816**
		One-error	0.6037	0.6847	**0.5772**	0.6287	0.5828
		Coverage	2.4051	2.7632	2.3328	2.4782	**2.3111**
		Ranking Loss	0.1611	0.1896	0.1561	0.1723	**0.1549**
		Average Precision	0.5811	0.5148	0.5942	0.5656	**0.5955**
Plant	**95**	Hamming Loss	0.0870	0.0886	**0.0848**	0.0883	0.0868
		One-error	0.6641	0.7253	**0.6256**	0.6667	0.6308
		Coverage	2.4231	2.7658	2.3615	2.3062	**2.2730**
		Ranking Loss	0.2110	0.2431	0.2053	0.2037	**0.1957**
		Average Precision	0.5365	0.4856	0.5604	0.5411	**0.5641**
Medical	890	Hamming Loss	0.0171	0.0277	0.0238	0.0237	**0.0165**
		One-error	0.2643	0.6934	0.4174	0.4028	**0.2637**
		Coverage	2.7237	6.4579	4.1051	3.7024	**2.9279**
		Ranking Loss	0.0425	0.1264	0.0683	0.0756	**0.0473**
		Average Precision	0.7957	0.4387	0.6739	0.6758	**0.7938**

4.2. Consideration of Parameters

In ML-ReliefF, we consider two parameters: the number of nearest neighbors K and the number of training samples m.

For K, we have mentioned that for high-dimensional data, it is sparse but one sample may still have a few similar neighbors. To maintain the integrity of the information of the data, we should select smaller values for K. As the Medical dataset is very sparse, we set the value of K on Medical to 1. On other datasets we test and validate the value of K from 1 to 5 as shown in Fig.3.

For m: as we know, when the number of the training samples increases, the average precision increases, but the computational time also increases. So one needs to find a balance for m to make both the performance and the time optimal. We tested m from 100 to 500 on the datasets besides the Medical and the results are shown in Fig.4.
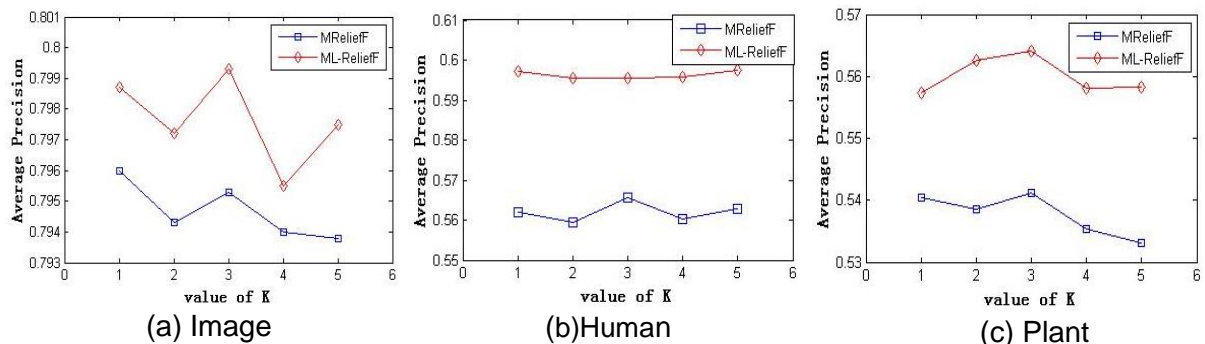

(a) Image   (b)Human   (c) Plant

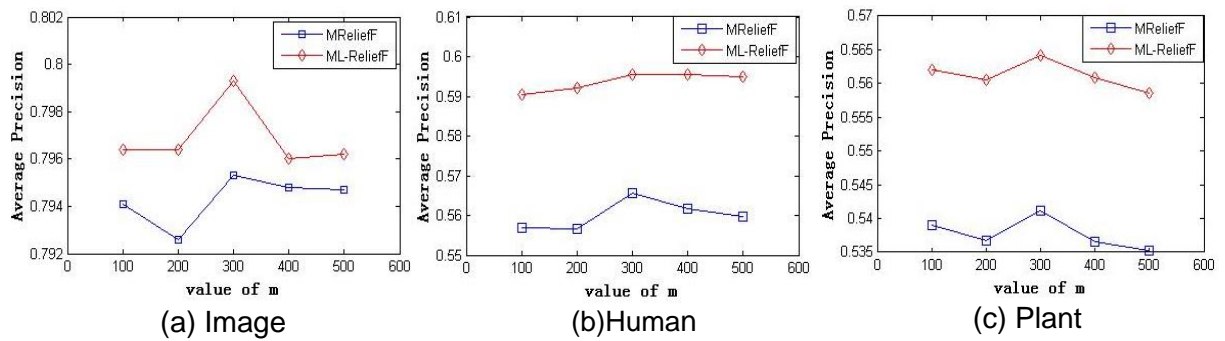**Figure 3. Comparisons of Different Values of K**

Figure 4. Comparisons of Different Values of m

From the experiment results in Fig.3 and Fig.4, we can see that for parameter K, both MReliefF and ML-ReliefF yield good performance at the value of 3. For parameter m, after the value reaches 300, the classification results become flat. So in our experiments, we set the parameter K to 3 and parameter m to 300.

All the results of the experiments demonstrate the effectiveness of the proposed method, with which, the selected features represent the most discriminant characteristics.

5. Conclusion and Future Work

In this paper, we present a new feature selection method for multi-label data called ML-ReliefF. The method introduces the concept of label set, modifies the calculation and regulation of the nearest neighbors and adds the similarities between the samples into the updating formula of feature weights. The proposed method works better than the MReliefF on most datasets when used with the ML-kNN classifier, in terms of the classification accuracy. The proposed ML-ReliefF can still be improved in several aspects. For instance, the correlations between labels can be further taken into account. The future work will focus on how to explore the label correlations into the multi-label feature selection.

Acknowledgments

References

[1] G. Tsoumakas and I. Katakis, "Multi-Label Classification: An Overview", International Journal of Data Warehousing and Mining, vol.3, no.3, (**2007**), pp.1-13.

[2] G. Tsoumakas, I. Katakis and I. Vlahavas, "Mining Multi-label Data. Data Mining and Knowledge Discovery Handbook", Part 6, Maimon O, Rokach L (Ed.), Springer, vol.2, (**2010**), pp.667-685.

[3] R. E. Schapire and Y. Singer, "Boostexter: a boosting-based system for text categorization", Machine Learning, vol.39, (**2000**), pp.135–168.

[4] S. Godbole and S. Sarawagi, "Discriminative methods for multi-labeled classification", In PAKDD '04: eighth Pacific-Asia conference on knowledge discovery and data mining, (**2004**); Berlin, Springer.

[5] J. Fürnkranz, E. Hüllermeier, E. L. Mencía and K. Brinker, "Multilabel classification via calibrated label ranking", Machine Learning, vol.73, no.2, (**2008**), pp.133-153.

[6] A. Clare and R. D. King, "Knowledge discovery in multi-label phenotype data", In: De Raedt L, Siebes A, eds. Lecture Notes in Computer Science 2168 .Berlin : Springer, (**2001**), pp.42-53.

[7] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification", In: Dietteroch T G, Bercker S, Ghahramani Z, eds. Advances in Neural Information Processing Systems 14.Cambridge, MA: MIT Press, (**2002**) pp.681-687.

[8] Z. Barutcuoglu, R. E. Schapire and O. G. Troyanskaya, "Hierarchical multi-label prediction of gene function", Bioinformatics, vol.22, no.7, (**2006**), pp.830-836.

[9] M. R. Boutell, J. Luo, X. Shen and C. M. Brown, "Learning multi-label scene classification", Pattern Recognition, vol.37, no.9, **(2004)**, pp.1757-1771.

[10] G. J. Qi, X. S. Hua, Y. Rui, J. Tang, T. Mei and H. J. Zhang, "Correlative multi-label video annotation", In: Proceedings of the 15th ACM International Conference on Multimedia. New York, NY: ACM Press, **(2007)**.

[11] C. G. M. Snoek, M. Worring, J. C. van Gemert, J. M. Geusebroek and A. W. M. Smeulders, "The challenge problem for automated detection of 101 semantic concepts in multimedia", In: Proceedings of the 14th ACM International Conference on Multimedia (ACM Multimedia'06), **(2006)**; Santa Barbara, CA.

[12] L. Tang, S. Rajan and V. K. Narayanan, "Large scale multi-label classification via metalabeler", In: Proceedings of the 19th International Conference on World Wide Web (WWW'09), **(2009)**; Madrid, Spain.

[13] B. Yang, J. T. Sun, T. Wang and Z. Chen, "Effective multi-label active learning for text classification", In: Proceedings of the 15thACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09), **(2009)**; Paris, France.

[14] G. Tsoumakas and I. Vlahavas, "Random k-labelsets: An ensemble method for multilabel classification", In: Kok J N, Koronacki J, de Mantaras R L, Matwin S, MladeničD, Skowron A, eds. Lecture Notes in Artificial Intelligence 4701, Berlin:Springer, **(2007)**, pp.406-417.

[15] M. L. Zhang and Z. H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning", Pattern Recognition., vol.40, no.7, **(2007)**, pp.2038-2048.

[16] M. L. Zhang, "ML-RBF: RBF neural networks for multi-label learning", Neural Processing Letters, vol.29, no.2, **(2009)**, pp.61-74.

[17] M. L. Zhang and K. Zhang, "Multi-label learning by exploiting label dependency", In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10), **(2010)**; Washington, D. C.

[18] H. Liu and H. Motoda, "Feature Selection for KnowledgeDiscovery and Data Mining", Springer **(1998)**.

[19] I. T. Jolliffe, "Principal Component Analysis", Berlin:Springer, **(1986)**.

[20] R. A. Fisher, "The use of multiple measurements in taxonomic problems", Annals of Eugenics, vol.7, no.2, **(1936)**, pp.179–188.

[21] K. Yu, S. Yu and V. Tresp, "Multi-label informed latent semantic indexing", In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'05), **(2005)**; Salvador, Brazil.

[22] Y. Zhang and Z. H. Zhou, "Multi-label dimensionality reduction via dependency maximization", ACM Transactions on Knowledge Discovery from Data. vol.4, no.3, **(2010)**, pp.14.

[23] S. Ji and J. Ye, "Linear dimensionality reduction for multi-label classification", In: Proceedings of the 21stInternational Joint Conference on Artificial Intelligence (IJCAI'09), **(2009)**; Pasadena, CA.

[24] B. Qian and I. Davidson, "Semi-supervised dimension reduction for multi-label classification", In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10), **(2010)**; Atlanta, GA.

[25] I. Kononenko, "Estimating attributes: Analysis and extensions of relief", Springer, **(1998)**, pp.171-182.

[26] M. L. Zhang, J. M. Peña and V. Robles, "Feature selection for multi-label naive bayes classification", Information Sciences, vol.179, no.19, **(2009)**, pp.3218-3229.

[27] D. Kong, C. H. Q. Ding, H. Huang and H. Zhao, "Multi-label reliefF and F-statistic feature selections for image annotation", In CVPR, **(2012)**, pp.2352–2359.

Authors
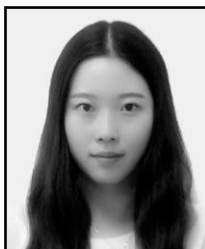
Yaping Cai, she is currently a M.S. candidate in the School of Computer Science and Technology, Nanjing Normal University, and received her B.S. degree from the same university in 2013. His research interests include machine learning, pattern recognition.

Ming Yang, he received his Ph.D. degree in the department of computer science and engineering from Southeast University at Nanjing in 2004. He received his M.S. degree in the department of mathematics from University of Science & Technology of China, and his B.S. degree in the department of mathematics from Anhui Normal University, in1990.

Yang Gao, he received his Ph.D. degree in computer science from Nanjing University in 2000. He received his M.S. degree in computer aided design from Nanjing University of Science and Technology in 1996, and his B.S. degree in Dalian University of Technology in 1993 .His research interests include Reinforcement Learning, Intelligent System, Image Process and Video Surveillance.
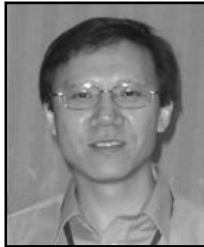
Hujun Yin, he received BEng and M.Sc. degrees from Southeast University and PhD degree from University of Yorkin 1983,1986 and1996, respectively. His research interests include neural networks, self-organising systems in particular, pattern recognition, bio-/neuro-informatics, and machine learning applications.