

## A reformative feature selection algorithm in fall detector application

Yuqi Cai<sup>1,a</sup>, Zhonghua Zhao<sup>2,b</sup>

<sup>1</sup>State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China

<sup>2</sup>National Computer Network Emergency Response Technical Team Coordination Center of China (CNCERT/CC), Beijing, 100029, China

<sup>a</sup>yqcai1991@163.com

**Keywords:** feature selection, fall detector, feature weight

**Abstract.** There are many fall detector applications on the Internet, all the applications aim to run more quickly and judge the status more accurate. When meeting masses of fall detector data, it needs to use lots of mathematical or geometrical features to judge the status, which is not benefit for the application. This paper is aimed to use a feature selection algorithm to calculate the most effective features, which is significant to the detect result, to reduce the cost of the feature selection process. Therefore, this paper has proposed an improved algorithm to advance the accuracy of the selection of the most important features, and then use different classify algorithm to classify with subsets after relief algorithm and reformative relief algorithm. The result shows that the reformative relief algorithm can provide a more effective subset which can reduce the feature size and improve the accuracy of classify samples.

### Introduction

With elder people's number grows bigger, the phenomenon of falling down is becoming more and more common in our daily life. Therefore, the need of detecting falling of elder people has been increased<sup>[1-4]</sup>. The main idea of solve this problem depends on feature selection algorithm, which includes filter algorithm and wrapper algorithm and it has some drawbacks such as the accuracy of the algorithm is not enough and the ability in solve different types of problems. In this paper we propose a new Relief algorithm which will use several different type samples and several same type samples to calculate the feature weight, which can improve the accuracy in fall detect circumstance.

Feature selection plays a key role in many pattern recognition problems such as image classification<sup>[5-6]</sup>, the main goal of feature selection is to decrease the dimensions of features without affect the result. Among feature selection algorithms, Relief algorithm is widely used by many scientists. Relief algorithm is proposed by Kira in 1992<sup>[7]</sup>, it is an algorithm based on feature weight. The main idea of Relief algorithm is give every feature a weight to present relation with different types, use Hypotheses-Margin to train the features in iteration<sup>[8]</sup>. Relief algorithm is able to verify those features which are not related in statistics and target attributes<sup>[9]</sup>.

Although Relief algorithm is widely used, it has some drawbacks as follows:

- (1) Select samples randomly will make small type of examples have less chance to be picked, sometimes it will be neglected completed. However, without small type of samples, the feature weight is not rational.
- (2) All kinds of samples may have effect in feature weight, when a type of examples has a larger number, the feature weight may be larger compared with other type of samples.

In this paper, we propose a reformative Relief algorithm which will use feature weight to improve the accuracy compared with Relief algorithm, then use several classify algorithm with original different subsets. Through the experiment, it can be proved that reformative Relief algorithm has improved 2.7% compared with full-set and improved 2.5% compared with subset after Relief algorithm. The feature size has reduced 71.4% compared with full-set.

## Our Reformative Algorithm

The main modified algorithm formula is as follows:

$$diff(A, X_R, Near\_Hit) = \frac{|X_R[A] - Near\_Hit[A]|}{\max(A) - \min(A)} \quad (1)$$

$$diff(A, X_R, Near\_Miss) = \frac{|X_R[A] - Near\_Miss[A]|}{\max(A) - \min(A)} \quad (2)$$

$$Sample\_Weight = \frac{\sum_{i=1}^k diff(A, X_R, Near\_Miss[i])}{\sum_{j=1}^k diff(A, X_R, Near\_Hit[j])} \quad (3)$$

$$W[A] = W[A] + \sum_{i=1}^m Sample\_Weight \quad (4)$$

Diff is to calculate the distance of current sample and the near sample, Sample\_Weight is the feature weight of the current sample, it will calculate the division of the sum of the nearest k samples of different type and the sum of the nearest k samples of same type. If the division is big, it means that the feature can distinguish the different type of sample more than the same type of sample; If the division is small, it means that the feature can distinguish the different type of sample less than the same type of sample.

The detail steps of the reformative algorithm are as follows:

The training dataset is D, we take samples from D for M times, the nearest sample is k, and the output is the feature weight.

Step1: Set all the feature weight zero

Step2: Take a sample R randomly from dataset D;

Step3: Take the nearest k samples which are same type of sample R and take the K samples which are the different type of sample R.

Step4: Calculate the feature weight of sample R

Step5: Do all these above for M times.

## Experimental results

**Data Source.** In order to make experiment more effectively, the data source of examples is more than 160,000. It contains the information as follows:

Person, location of fallen detective equipment, timestamp, the accelerate of X axis, the accelerate of Y axis, the accelerate of Z axis, the square root of X accelerate and Y accelerate, the square root of X accelerate and Z accelerate, the square root of Y accelerate and Z accelerate, the square root of X, Y, Z accelerate, status (it contains 9 status of the person).

Table 1 Data source of experiment

Data type	Property type	Examples	Properties
Various of types	Numeric	160,000	7

**Experiment method.** The experiment is executed by different person and different location of falling detective equipment. And the basic information is as follows:

Table 2 Basic information of falling down status

Person name	Location of detect equipment	Status of person
A01-A05, B01-B05, C01-C05, D01-D05, E01-E05	Ankle-left, Ankle-right, chest, belt	walking, sitting down, sitting, standing up from sitting, lying, standing from lying, lying down, sitting on the ground, falling

Each time we will only use the data of one person and one location to select the most significant features to distinguish the status of falling and other status, which may decrease the dimension of feature space.

The experiment software we used is weka<sup>[10]</sup>, which is compiled by Java for machine learning and data mining, it contains visual tools and algorithms of all kinds of data analysis and predict model and provide a graph user interface for users to use conveniently

**Experiment analysis.** In this experiment, there are 7 parameters to determine the final result, the parameters are as follows: x axis accelerate, y axis accelerate, z axis accelerate, square root of xy, square root of xz, square root of yz, square root of xyz.

First, we will use weka to get feature subsets of Relief algorithm and Reformative algorithm. The results are as follows:

```
0.089    4 y
0.0684   6 sqrt(xy)
0.061    3 x
0.0578   7 sqrt(xz)
0.0486   8 sqrt(yz)
0.0461   9 sqrt(xyz)
0        5 z
```

```
0.1118   7 sqrt(xz)
0.1045   3 x
0.0816   6 sqrt(xy)
0.0766   9 sqrt(xyz)
0.0764   4 y
0.0563   8 sqrt(yz)
0.0544   5 z
```

Fig.1 Feature weight after relief algorithm

Fig.2 Feature weight after reformative algorithm

In the Fig.1, we can know that the feature sequence is [4,6,3,7,8,9,5], the feature weight drops down apparently from feature 4 to feature 6, so the feature subset of Relief is feature 4. In the Fig.2, we can know that the feature sequence is [7,3,6,9,4,8,5], the feature weight drops down obviously from feature 3 to feature 6, so the feature subset of reformative algorithm is feature 7 and feature 3.

Table 3 Subset after relief and reformative relief

Feature Selection algorithm	Subset size	Subset
Relief	1	[4]
Reformative Relief	2	[3,7]

According to Table 3, we will use some common classify algorithm, in order to get the classify accuracy, the result is as follows:

Table 4 Classify accuracy with different subset

Classify algorithm	Full-set	Subset after relief	Subset after reformative relief
SVM	93.8%	93.8%	96.1%
LDA	93.4%	93.8%	94.9%
Bayes	90.2%	92.9%	95.7%
Adaboost	94.1%	91.8%	95.7%

From Table 4, we can know that the accuracy of classify algorithm with subset after reformative Relief algorithm has improved 2.7% compared with full-set and improved 2.5% compared with subset after Relief algorithm. The feature size has reduced 71.4% compared with full-set.

## Conclusion

In this paper, we have presented a reformative Relief algorithm. Our goal is to minimize the feature size, which is used in the Android application of fall detection. According to the experiment, we can discover that reformative Relief algorithm has improved the classify accuracy more and reduce the feature size obviously.

## Acknowledgements

This work is supported by NSFC (Grant Nos. 61300181, 61502044), the Fundamental Research Funds for the Central Universities (Grant No. 2015RC23).

## References

- [1] Doughty K, Lewis R and et al, The Design of a Practical and Reliable Fall Detector for Community and Institutional Telecare [J]. Journal of TeleMedicine and Telecare, 2000:p. 150-154.
- [2] Rubenstein L Z, Falls in Older People Epidemiology, Risk Factors and Strategies for Prevention [J].Age and Ageing, 2006:p.ii37~ii41.
- [3] Fuller, G. F., Falls in the Elderly [J].American Family Physician, 2000:p.2159~2168.
- [4] M. Mubashir, L. Shao and L. Seed, A survey on fall detection: Principles and approaches. Neurocomputing, 2013. 100(0): p>144-152.
- [5] Guyon, I. And Elisseeff, A. (2003) An Introduction to Variable and Feature Selection. The Journal of Machine Learning Research, 3, 1157-1182.
- [6] Song, D. J. And Tao, D. C. (2010) Biologically Inspired Feature Manifold for Scene Classification, IEEE Transaction on Image Processing, 19, 174-184.
- [7] Marko R S, Kononenko I. Comprehensible interpretation of relief estimates[C]//In: Brodley C E, Danyluk AP, Eds. Proceedings of the 18th International Conference on Machine Learning. San Francisco: Morgan Kaufmann,2001:440-443
- [8] Kononenko I. Estimation Attributes: Analysis and Extensions of Relief[C]//Proceedings of the 1994 European Conference on Machine Learning, 1994:171-182.
- [9] T.T Nguyen, Z. Li, T. Silander, et.al. Online Feature Selection for Model-based Reinforcement Learning, in:ICML, 2013, pp. 1-9
- [10]Witten I H, Frank E. Data Mining: Practical Machine Learning Tools and Techniques[J]. Biomedical Engineering Online, 2011, 5:51(1):95–97..