A New Dynamic Method of Machine Learning From Transition Examples

Xiao-dan Zhang[1], De-gan Zhang [2, 3,]*, De-xin Zhao[2, 3], Xue-Jing Kang[2, 3] , Xiao-dong Qiao[1]

[1]Institute of Scientific and Technical Information of China, Beijing, 100038, China
[2]Tianjin Key Lab of Intelligent Computing & Novel software Technology, Tianjin University of Technology, China
[3]Key Laboratory of Computer Vision and System (Tianjin University of Technology), Ministry of Education, China
E-mail: zhangshenyang@126.com, gandegande@126.com, *Corresponding author

***Abstract*—It's well known machine learning from examples is an effective method to solve non-linear classification problem. A new dynamic method of machine learning from transition example is given in this paper. This method can improve the traditional method ID3 which learns from static eigenvalues of examples. The limits of the traditional method ID3 lie on no comprehension and no memory, especially, no the varieties and dynamic correlation of eigenvalues. In the new method, it can learn from dynamic eigenvalues, the change of data can be learned because the training data is the initial eigenvalue and the end eigenvalue in the interval. All eigenvalue's varieties and correlation can be understood and remembered in application. By test experiments, the new method can be used as classifier when the multi-parameters are dynamic correlation, and it has special use in the many kinds of information fusion fields.**

***Index Terms*—machine learning, entropy, dynamic, classifier, example**

I. INTRODUCTION

Machine learning from examples is an effective method to solve non-linear classification problem[1]. Generally, data identification and classification decides by three kind of information as follows: data values, regular of change, data dependency—in particular, studying the dynamic correlation is very difficult to use to describe traditional mathematical models, because of the non-linear, non-deterministic and its illegibility [2]. Recently, scholars have pushed off many methods in the field of machine learning from examples, neural network machine learning from examples bear the brunt because of its flexible and simple, but it was limited by the astirngency of large-scale problems and the minimum of partial problems [3]; Researching on the effect and processing power of other methods [4] wants to dig out the hidden important information in the dynamic changing data. The ID3 (Interactive Dichotomic version3) method of J. R. Quinlan is a static characteristic value for the study, which is very influential in the international community [5-7]. In this paper, we pushed off the MLBET (machine learning Based on Eigen-value Transition) which based on the ID3 method has the ability to machine learning eigenvalue change. The method accepts eigenvalue transition and classification of an input sample for a certain interval, the machine learning process can integrate to understand and remember the above-mentioned three kinds of information and increase the information storage capacity of decision tree to improve classification capability [8-9].

Manuscript received May 1, 2010; revised January 1, 2011; accepted March 1, 2011.
*corresponding author.

II. THE PRINCIPLE OF MLBET METHOD

Compared to the traditional method ID3, the most advancement of MLBET is the increasing of accepting and remembering data information. Each example accepted by MLBET contains two eigenvalues which designated for the initial and end eigenvalue in the interval. It can learn information from dynamic correlation. Each example is decomposed into a number of pairs of dual sequence features in order to ensure a balanced number of each eigenvalue. To eigen-decomposition make features dual quadruple; According to information theory [10-12], the feature of largest mutual information will be the most important factor of ability to make judgment in the study. The data is divided into multiple sub-sets, each sub-set is also divided by the most ability to make judgment until the data of sub-set is in the same kind. Doing this can get a decision tree, further training make the decision tree to be continually revised until the end of the training. When MLBET dealing with classification problems, besides based on the characteristic values, it can also make comprehensive judgment based on feature changing information in certain interval and three factors of dynamic correlation.

III. DEFINITION

The first procedure of MLBET method is multi-valued eigen-decomposition to get sequence characteristics of four values. Process of Machine learning has always been dealing with sequence characteristics of dual process.

A. Define example information

Definition 1. Eigenvalue transition—eigenvalue changes from the low value to high value or high value to low value in specified interval.

Definition 2. Set an example set S={ S_1, S_2, . . . , S_e}, the total number of example $|S|=e$, each

doi:10.4304/jsw.6.10.2064-2067

example S_i is $S_i =(S'_i , S''_i , C_f)$ in which S'_i , S''_i is two sets of eigenvalue in specified interval, S'_i is the initial value, S''_I is the end value, S contains g classes-- $U_1, U_2, \ldots U_g$ which repesents $C_1, C_2, \ldots, C_g$ g classes separately, if eigenvalue S'_i becomes S''_i , eigenvalue S_i belongs to class C_f.

Definition 3.There are h eigenvalues, namely $A_1, A_2, \ldots, A_h$, the *jth* eigenvalue A_j has $m(j)$ values, namely: $(a_{j1}, a_{j2}, \ldots, a_{jm(j)})$.

Definition 4. There is $S'_i=(s'_{i1}, s'_{i2}, \ldots, s'_{ih})$for S'_i and $S''_i=(s''_{i1}, s''_{i2}, \ldots, s''_{ih})$; then $(s'_{i1}, s'_{i2}, \ldots, s'_{ih})and(s''_{i1}, s''_{i2}, \ldots, s''_{ih})$is the counterpart on the $(A_1, A_2, \ldots, A_h)$, so eigenvalue s'_{ij} and s''_{ij} has $m(j)$ values.

B. Eigen-decomposition and encode

Definition 5.Be dual eigenvalue, correspondingly, s'_{ij} and s''_{ij} changes from multi-valued features to two-valued feature. The eigenvalue of decomposition should be $(x'_{ij1}, x'_{ij2}, \ldots, x'_{ijm(j)})$ and $(x''_{ij1}, x''_{ij2}, \ldots, x''_{ijm(j)})$, each eigenvalue encode for 0 or 1, presents by u_0 and u_1 . Accordingly, the number of eigenvalue changes from h originally to $\sum_{j=1}^{h} m(j)$.

Definition 6.Be quadruple of sequence characteristics, each two value of eigenvalue based on interval can be expressed a sequence characteristic $<x'_{ijk}, x''_{ijk}>$ of which value can take from $<0, 0>$, $<0, 1>$, $<1, 0>$, $<1, 1>$, separately remembers as V_1 , V_2 , V_3 , V_4 . Finally, eigenvalue S_I can be expressed $S_i=(<x'_{ijk}, x''_{ijk}>(j=1, 2, \ldots, h;\ k=1, 2, \ldots, m(j)), C_f)$.

IV. CALCULATION OF MUTUAL INFORMATION

According to definition 2, the contingency of U_b happens:

$$P(U_b)=|U_b|/|S| \quad (1)$$

and $\sum_{b=1}^{g} P(U_b)=1$

The conditional probability of example sub-set X_{bd} which the value of sequence characteristic $<x'_{ijk}, x''_{ijk}>$ gets V_d in class U_b is

$$P(V_d \mid U_b)=|X_{bd}|/|U_b| \quad (2)$$

and $\sum_{d=1}^{4} P(V_d|U_b)=1$

The probability of example sub-set F_d which gets value V_d in sequence characteristic $<x'_{ijk}, x''_{ijk}>$ is

$$P(V_d)=|F_d|/|S| \quad (3)$$

and $\sum_{d=1}^{4} P(V_d)=1$

The probability of example set F_{bd} which belongs to U_b class and gets value V_d in sequence characteristic $<x'_{ijk}, x''_{ijk}>$ is

$$P(U_b|V_d)=|F_{bd}|/|F_d| \quad (4)$$

and $\sum_{d=1}^{4} P(U_b|V_d)=1$

The mathematical model of information source which consists of the probability $P(U_i)$ of information $U_b(b=1, 2, \ldots g)$happens is

$$[U, P]=\begin{bmatrix} U_1, U_2, \ldots, U_g \\ P(U_1), P(U_2), \ldots P(U_g) \end{bmatrix} \quad (5)$$

Correspondingly, the self information of information U_i is

$$I(U_b)=log\frac{1}{P(U_b)} \quad (6)$$

According to the definition of information entropy and formula (6), thrusting out:

$$H(U)=\sum_{b=1}^{g} P(U_b)\, I(U_b) = \sum_{b=1}^{g} P(U_b) log\frac{1}{P(U_b)} \quad (7)$$

According to the definition of back proving entropy and formula (7), the back proving entropy of sequence characteristic $<x'_{ijk}, x''_{ijk}>$ getting value V_d is

$$H_{ijk}(U|V_d)=\sum_{b=1}^{g} P(U_b|V_d)\, log\frac{1}{P(U_b|V_d)} \quad (8)$$

According to formula (8), the counterpart condition entropy is

$$H_{ijk}(U|V)=\sum_{d=1}^{4} P(V_d) \sum_{b=1}^{g} P(U_b|V_d)\, log\frac{1}{P(U_b|V_d)}$$
(9)

According to the definition of self information and formula (7) and (9), the average of self information of sequence characteristic is

$$I_{ijk}(U|V)= H(U)- H_{ijk}(U|V) = \sum_{b=1}^{g} P(U_b) log\frac{1}{P(U_b)} - \sum_{b=1}^{g} \sum_{d=1}^{4} P(V_d)P(U_b|V_d)\, log\frac{1}{P(U_b|V_d)} \quad (10)$$

V. Algorithm of Machine Learning

All the machine learning process consists of two parts: the decision tree generated by the basic example sets and the decision tree is constantly modified by the Machine learning of the test set data.

A. Main algorithm

(1)The example set $S=\{S0+S1\}$ in which $S0$ is basic set, $S1$ is testing set.

(2)Select $S0$, call tree-building algorithm

(3)If $S1 \neq \{\Phi\}$, take $S_i \in S1$ arbitrarily, there are $S1=S1-S_i$, $S_i=(<x'_{ijk}, x''_{ijk}>(j=1, 2, \ldots, h; k=1, 2, \ldots, m(j)), C_f)$. According to the decision tree of step (2), it is identified as C'_f class. If $C_f = C'_f$, go back to step (3); If $C_f \neq C'_f$, $S0=S0+S_I$ and go back to step (2).

B. Tree-building algorithm

(1)All of the sequence characteristic $<x'_{ijk}, x''_{ijk}>$ in current sub-sets, according to formula (7), (9), (10)gets self information set $I=\{I_{ijk}(U, |V) (<x'_{ijk}, x''_{ijk}>(i=1, 2, \ldots, n; j=1, 2, \ldots, h; k=1, 2, \ldots, m(j))\}$, take $I_{max}=max\{ I_{ijk}(U, |V)\}$ so the counterpart sequence characteristic remember as c.

(2)Get four sub-sets:

$W_P=\{ S_i | S_i \in S1$ AND the value of $<x'_{max}, x''_{max}>$ is V_P }

$P=1, 2, 3, 4$

(3)If $\exists S_{i1}, S_{i2} \in W_P$, S_{i1} belongs to C_{f1} class, S_{i2} belongs to C_{f2} class, moreover $C_{f1} \neq C_{f2}$, go back to (1); if $\forall S_{i1}, S_{i2} \in W_P$, S_{i1} belongs to C_{f1} class, S_{i2} belongs to C_{f2} class, moreover $C_{f1}= C_{f2}$, the corresponding branch is C_{f1} class and go back to the calling.

VI. Test of MLBET and ID3 method

A. Example training

MLBET and ID3 methods were used separately for the condition identification of 4#, 10# main transformer in Hydropower Station. It can be used as Identifying three states of device: normal, dangerous and breakdown. The problem has non-linear, non-deterministic and illegibility apparently.

Because of the limit of the static eigenvalue machine learning method, such as three characteristic values: power, gas and temperature below the warning value, if the static analysis would suggest that it is absolutely safe. In fact, if power and the temperature of main transformer continued to drop over time, while the gas continued to rise, even though the value is not high, but it also shows that the main transformer may partially breakdown.

MLBET method can accurately judge this condition. Featured example S=1500, A_1 =power, A_2=gas, A_3=temperature; according to the division of human experts, three characteristic values are: low, low, medium, high, high; the eigenvalue is 15 after decomposition, the example number of universal set is 4^{15}; the interval take $\Delta t= t- t'$, each eigenvalue compose a pair of dual sequence according to the value taking on t, t' time. Δt=K is a constant of time interval, Δt=0 expresses acceptance of the static characteristic value; C_1 =normal, C_2 =dangerous, C_3 =breakdown. Take $S0$=350, $S1$=1150, learn according to algorithm.

This paper take 650 static eigenvalues (Δt=0) example as public set S_p and 950 dynamic eigenvalues (Δt=K)example S_L composed 1500 example sets of MLBET; also take 950 static eigenvalue (Δt=0)example S_i composed 1500 example sets of ID3; such as table 1:

TABLE 1
EXAMPLE SETS CROSS TABULATION IN TWO METHODS

Method	Class	Basic Set |S0|=350		Testing Set|S1|=1150	
		|**S**p|=150	|**S**i/ **S**L |=200	|**S**p|=500	|**S**i/ **S**L|=750
ID3 method |S|=1500	C_1	38	41	145	158
	C_2	43	65	166	283
	C_3	69	94	189	309
MLBET method |S|=1500	C_1	38	48	145	185
	C_2	43	56	166	288
	C_3	69	96	189	277

B. Contradistinction of testing application

It can be used separately in large-scale operating simulation machine of Hydropower Station after training.

Online diagnosis on the 4#, 10# main transformer (parameters identical), the contrasting condition in table 2:

C. Performance evaluation

The performance evaluation of two methods is as follows:

(1)Two methods can give the right judgment when the state is normal and the data is regular comparatively.

(2)In table 2, two methods can find the dangerous and breakdown when the parameter exceeds standard. For the dangerous and breakdown information is implicit in the data when the parameter below warning level, only MLBET method can give judgment. MLBET is dynamic sensitive.

(3)The breakdown which ID3 finds is in a distinct period, MLBET could find breakdown which is in the incubation period or early period. MLBET finds breakdown earlier.

TABLE 2
CROSS TABULATION OF TESTING APPLICATION EFFECT

	Work condition	*Class*	Example	Correct judge	Error judge	Correct rate
ID3 method	Empty carry 300	C_1	47	44	3	93. 6%
		C_2	136	94	52	69. 1%
		C_3	117	58	59	49. 6%
	Carry 1200	C_1	286	252	34	88. 1%
		C_2	434	242	92	55. 8%
		C_3	480	165	315	34. 4%
MLBET method	Empty carry 300	C_1	51	47	4	92. 2%
		C_2	126	112	14	88. 9%
		C_3	123	108	15	87. 8%
	Carry 1200	C_1	263	241	22	91. 6%
		C_2	431	358	73	83. 1%
		C_3	506	436	70	86. 2%

VII. CONCLUSION

Based on our tests of MLBET and ID3 method, we can draw the conclusion as follows:

(1)The method of this paper can be used to dynamic eigenvalue and as classifier in the dynamic correlation problem. It is proved that this method improves the exact rate than ID3 method by the comparison of testing application.

(2)The method of this paper can understand and remember dynamic information, especially the dynamic correlation information. The information storage of decision tree is increasing.

(3)The example not only can take the same interval but also can take different interval. It doesn't affect the machine learning effect. When the interval is 0, this method accepts the data as eigenvalue. So this method contains ID3 method.

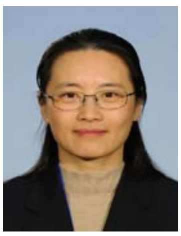
(4)The method of this paper has a fairly good stability for quite large-scale machine learning problems.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (No.60773073 & 61001174), Program for New Century Excellent Talents in University of China (No. NCET-09-0895), Tianjin Natural Science Foundation (No. 10JCYBJC00500), Key project of Ministry of Education of China (No.208010), Science and Technology Development Foundation of Tianjin Higher Education Institution (No. 20080805).

REFERENCES

[1] Limi Fu . Machine learning in Certainty-Factor-Based Multilonger Neural Network for Classification. IEEE Transaction on Neural Network , Jun 2008 Vol 19 No. 1 151~153
[2] Miroslav Kubat. Decision Trees Can Initialize Radical_Basis Function Networks . IEEE Transaction of Neural Networks , May 2008, Vol 19 No. 5 815~818
[3] Xu Li, Zhu Jing, Jiang Jing-ping. A Symmetric Ternary Logic Based Multi-valued Machine learning Network. Chinese journal of computers, 1998 Vol 21 No. 6 553~555
[4] Zhang Nai-yao. Neural Network and Fuzzy Control. Beijing: Tsinghua University Press, 2008, 10~80
[5] Li Jiao-cheng. Neural Network System Theory. Xi'an: Xi'an Electronic Science and Technology University Press, Sep 2006, 242~251
[6] Prasad. Tadepalli , Dokyeong. Model-based average reward reinforcement machine learning. Artificial Intelligence, Jun 2008 Vol 105 No. 1, 215~223
[7] Yishai. A Doron. A Portability by translation: A large-scaled study. Artificial Intelligence, 2009 Vol 117 No. 1, 210~212
[8] Roni Khardon. Machine learning strategies for planning domains. Artificial Intellegence, Sep 2009 Vol 123 No. 1/2 125~148
[9] Chen Bin, Hong Cheng-rong, Wang Ya-dong. Machine learning from examples of the maximum load problems and algorithms. Chinese journal of computers, Feb 2007 Vol 30 No. 2 133~138
[10] Cheng Ji-hua, Shi Peng-fei. Fast Mining Multiple-level Association Rules. Chinese journal of computers, 2008 Vol. 31 No. 11 1037~1041.
[11] Zhang Xiao-dan. A system of file Automated classification [P],china: 2010.12.
[12] Chen Wen-wei. Intelligent Decision Technology. Beijing: Electronics Industry Press, 2006,156~163



Xiao-dan Zhang Born in 1975, PHD. Graduated from Northeastern University, China. She is a researcher of Institute of Scientific and Technical Information of China, Beijing, 100038, China. Her research interest includes information mining, information fusion, AI or Web-based service, etc. Her E-mail: zhangshenyang@126.com.

De-gan Zhang Born in 1969, Ph.D., He is a researcher of Tianjin Key Lab of Intelligent Computing & Novel software Technology, Tianjin University of Technology, Tianjin, 300381, China. gandegande@126.com.

De-xin Zhao, Xue-Jing Kang and **Xiao-dong Qiao** are the colleagues of Xiao-dan Zhang and De-gan Zhang.