A Hybrid Approach of Neural Network and Memory-Based Learning to Data Mining

by

Chung-Kwan Shin, Ui Tak Yun, Huy Kang Kim, and Sang Chan Park

IEEE Transactions on Neural Networks, Vol. 11, No. 3, MAY 2000.

Review by: Thimaporn Phetkaew

Hybrid Approach to Data Mining

- Introduction
- Hybrid system of neural network and memory-based learning
- Experimental results
- Conclusion, limitations and future work

Introduction

- The knowledge representation of neural network (NN) is unreadable to humans.
- Memory-based reasoning (MBR) suffers from the feature-weighting problem.
- The k-nearest neighbor (k-NN) method --> all of the features presented are equally important.

Introduction

- In dynamic situations, the on-line learning property is crucial.
- Hybrid system is designed to take full advantage of the vast amount of memory.

Hybrid system of NN and memory-based learning

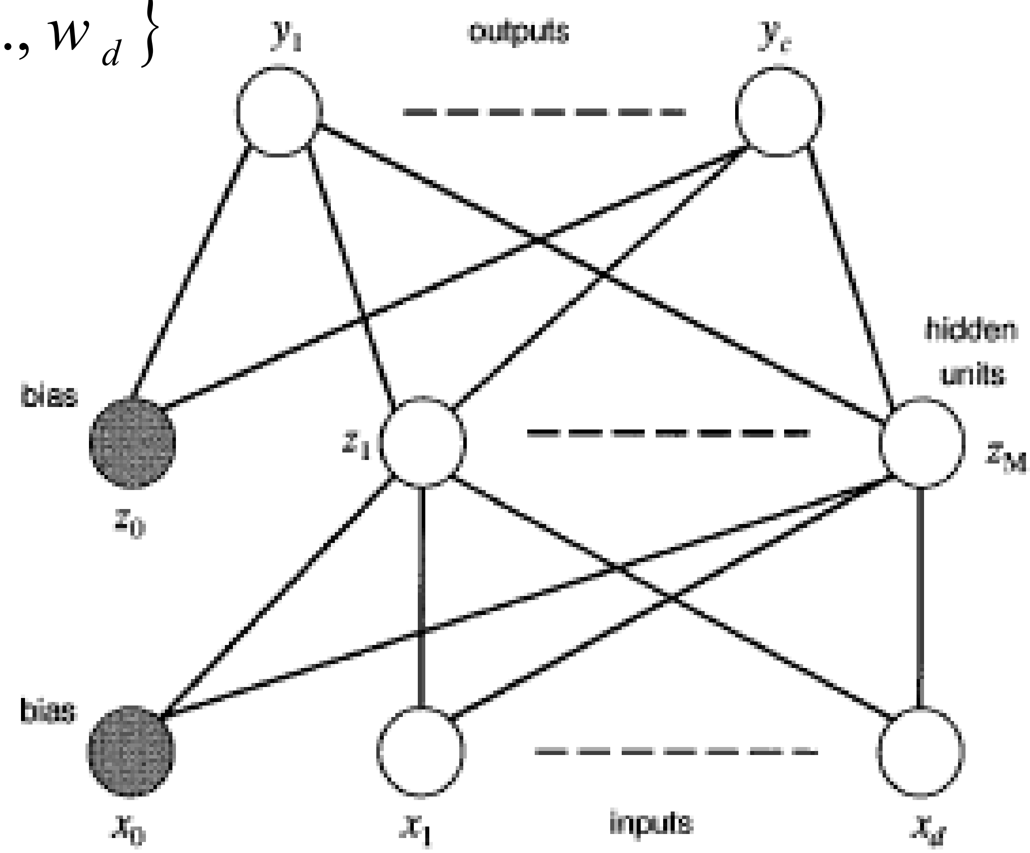
Information Source

Measure		Signal	Structure
	Intensity	I. Sensitivity	III. Saliency
	Variance	II. Activity	IV. Relevance

Framework of extracting feature weight from a trained NN

Hybrid system of NN and memory-based learning

$$\{w_1, w_2, \ldots, w_d\}$$



$$b_k = \sum_{j=0}^{M} w_{kj}^{(2)} z_j$$

$$a_j = \sum_{i=0}^{d} w_{ji}^{(1)} x_i$$

Example of fully connected network with one hidden layer

Hybrid system of NN and memory-based learning

■ Feature weighting algorithms using a trained NN

- Sensitivity --> input feature x_i

$$S_i = \frac{\left(\sum L \frac{\left| P^0 - P^i \right|}{P^0} \right)}{n}$$

Hybrid system of NN and memory-based learning

■ Feature weighting algorithms using a trained NN

- Activity --> a hidden node z_i

$$A_j = \left(w_j^{(2)}\right)^2 \cdot \mathrm{var}\left(g\left(\sum_{i=0}^{d} w_{ji}^{(1)} x_i\right)\right)$$

- --> an input node x_i

$$A_i = \sum_{j=1}^{M}\left(\left(w_{ji}^{(1)}\right)^2 \cdot A_j\right)$$

Hybrid system of NN and memory-based learning

■ Feature weighting algorithms using a trained NN

- Saliency --> an input node

$$Saliency_i = \sum_{j=1}^{M} \left(\left(w_{ji}^{(1)} \right)^2 \cdot \left(w_j^{(2)} \right)^2 \right)$$

Hybrid system of NN and memory-based learning

■ Feature weighting algorithms using a trained NN

- Relevance --> a hidden node z_j

$$R_j = \left(w_j^{(2)}\right)^2 \cdot \mathrm{var}\left(w_{ji}^{(1)}\right)$$

- --> overall relevance of input node x_i

$$R_i = \left(\left(w_{ji}^{(1)}\right)^2 \cdot R_j\right)$$

Hybrid system of NN and memory-based learning

■ MBR with weighted features (k-NN method)

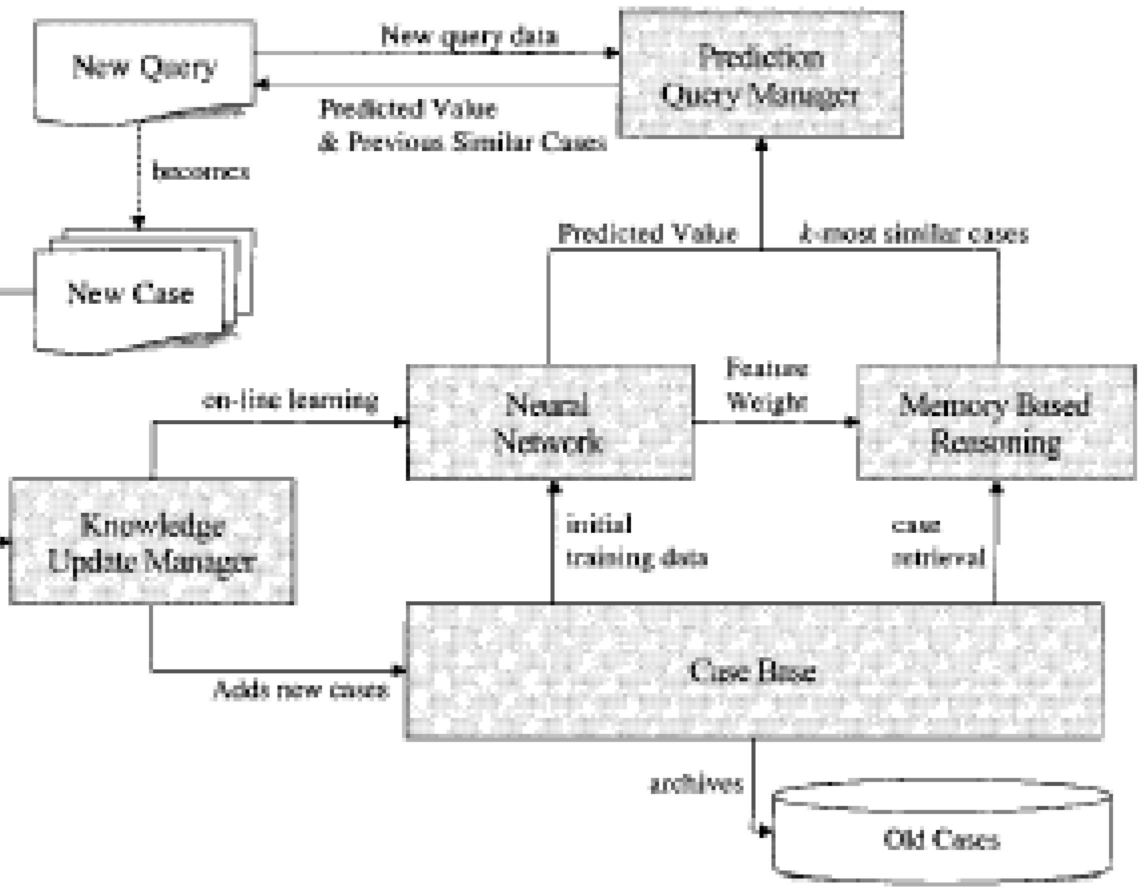
- Case $x = \{x_1, x_2, \ldots, x_n, x_c\}$
- Distance

$$\text{Distance}(x, q) = \sqrt{\sum_{f=1}^{n} W_f \times \text{difference}(x_f, q_f)^2}$$

$$\text{difference}(x_f, q_f) = \begin{cases} |x_f - q_f|, & \text{if feature } f \text{ is numeric} \\ 0, & \text{if feature } f \text{ is symbolic and } x_f = q_f \\ 1, & \text{otherwise} \end{cases}$$

Hybrid system of NN and memory-based learning

■ Integration of memory and NN-based learning



Hybrid system of NN and memory-based learning

- Integration of memory and NN-based learning
  - The rule of PQM rejection --> classification task
    - If P_M <> P_{NN} Then "Reject to answer"
      Else "Answer with P_M (or, identically, P_{NN})"
  - --> regression task
    - If $|P_M - P_{NN}| >= \varepsilon$ Then "Reject to answer"
      Else "Answer with average of P_M and P_{NN}"

Hybrid system of NN and memory-based learning

- Integration of memory and NN-based learning
  - Rejection ratio

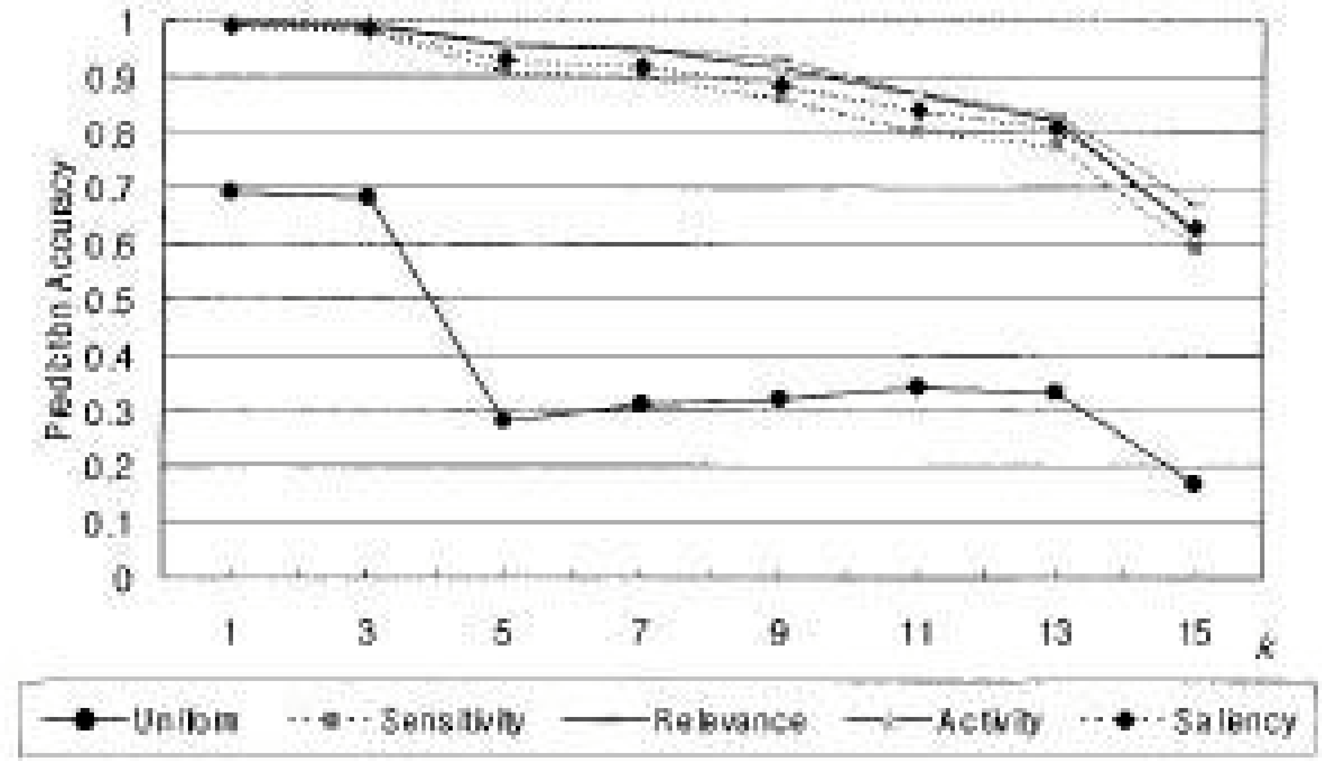
$$\text{Rejection ratio} = \frac{\text{Number of unanswered queries}}{\text{Total number of queries}}$$

Experimental results

TABLE I
DATASETS USED IN THE EXPERIMENTS AND THEIR NEURAL NETWORK LEARNING SETTINGS

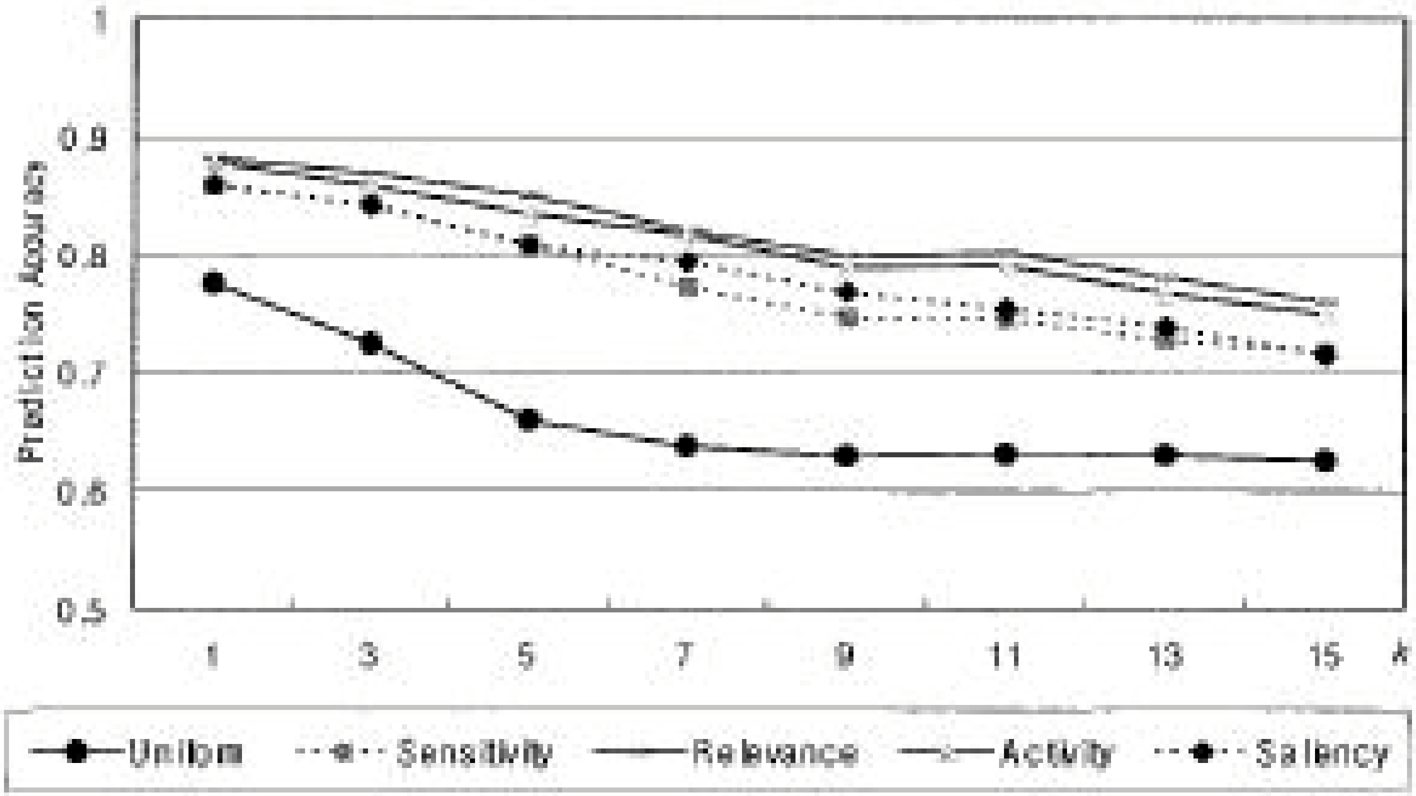
datasets used for the experiments						Neural Network Settings		
Dataset	Instances		Attributes		Output Classes	# of Hidden Nodes	Mean Iterations	Prediction Error
	Learning	Testing	Original	Random				
Odd Parity	350	150	4	7	2	6	727± 1528	0.0 ± 0.0
Sinusoidal	350	150	2	0	2	18	8442±1163	35.5±6.3
WDBC	468	101	30	0	2	20	3017± 1495	3.7 ± 0.0
WDBC+15				15			2826± 2221	5.7 ± 0.0
WDBC+30				30			3284± 814	5.9 ± 0.0
Credit	500	153	43	0	2	5	8820± 1454	20.20 ± 3.86
Credit+20				20		8	7553± 2143	22.43 ± 2.45
Credit+43				43		20	6870± 1915	24.2 ± 3.17
Sonar	104	104	60	0	2	6	6230±1565	16.5±2.15
Sonar+30				30		8	5995±2158	26.2±2.04
Sonar+60				60		9	8159±1927	27.1±3.54
MPG	300	92	9	0	real	9	5709± 2902	3.11 ± 0.41
MPG+4				4			5764± 3432	3.42 ± 0.52
MPG+9				9			3278± 2530	3.73 ± 0.32

Experimental results



Classification accuracy of the feature weighting method for the odd-parity problem

Experimental results



Classification accuracy of the feature weighting method for the sinusoidal problem

Experimental results

TABLE II
ODD PARITY PROBLEM (a) MEAN WEIGHTS OF RELEVANT FEATURES ($I_1 \ldots I_4$) AND IRRELEVANT FEATURES ($I_5 \ldots I_{11}$) AND (b) MEAN ERROR OF FEATURE WEIGHTING ALGORITHMS

	Sensitivity	Activity	Saliency	Relevance
I1 - I4	1.91± 0.60	2.06± 0.56	2.06± 0.42	2.04± 0.41
I5 - I11	0.48± 0.46	0.40± 0.60	0.39± 0.58	0.40± 0.62

(a)

k	Uniform	Sensitivity	Activity	Saliency	Relevance
1	30.8±3.4	0.9±1.3	0.8±1.4	1.0±1.8	0.2±0.4
3	31.7±2.6	1.8±2.5	1.3±2.3	1.7±2.7	0.7±1.2
5	72.0±1.5	9.3±11.1	4.5±6.7	6.9±9.5	4.3±5.5
7	69.0±1.0	10.1±12.2	4.8±7.0	8.3±11.4	5.2±7.0
9	67.9±2.6	14.1±15.0	6.9±7.3	11.5±12.7	8.4±8.3
11	65.8±2.6	19.7±15.5	13.1±9.7	16.3±13.0	13.4±10.2
13	66.7±3.1	22.1±16.6	16.7±9.1	19.1±11.2	17.9±9.3
15	83.1±2.0	41.1±21.3	33.1±13.7	37.3±11.0	38.1±10.4

(b)

Experimental results

TABLE III

SINUSOIDAL PROBLEM (a) MEAN WEIGHTS OF RELEVANT FEATURE (f_1) AND LESS RELEVANT FEATURE (f_2) AND (b) MEAN ERROR OF FEATURE WEIGHTING ALGORITHMS

	Sensitivity	Activity	Saliency	Relevance
f1	1.68±0.25	1.84±0.09	1.85±0.10	1.75±0.14
f2	0.32±0.25	0.16±0.09	0.15±0.10	0.25±0.14

(a)

k	Uniform	Sensitivity	Activity	Saliency	Relevance
1	22.4±2.2	14.2±4.4	11.7±1.3	12.1±1.3	14.0±1.7
3	27.6±3.5	15.8±6.1	13.1±2.9	14.2±3.7	15.9±3.2
5	34.0±4.9	19.1±6.7	15.0±3.7	16.5±3.8	19.2±4.5
7	36.3±4.3	22.8±8.2	17.9±2.6	18.3±4.3	20.7±5.7
9	37.1±4.9	25.3±8.4	20.1±3.7	21.1±4.0	23.2±6.3
11	36.9±5.0	25.5±9.5	19.9±3.2	20.9±4.5	24.6±5.2
13	37.0±3.5	27.3±9.0	21.9±4.1	23.3±5.1	26.1±6.6
15	37.5±3.9	28.3±8.6	24.0±5.3	25.1±6.5	28.5±7.1

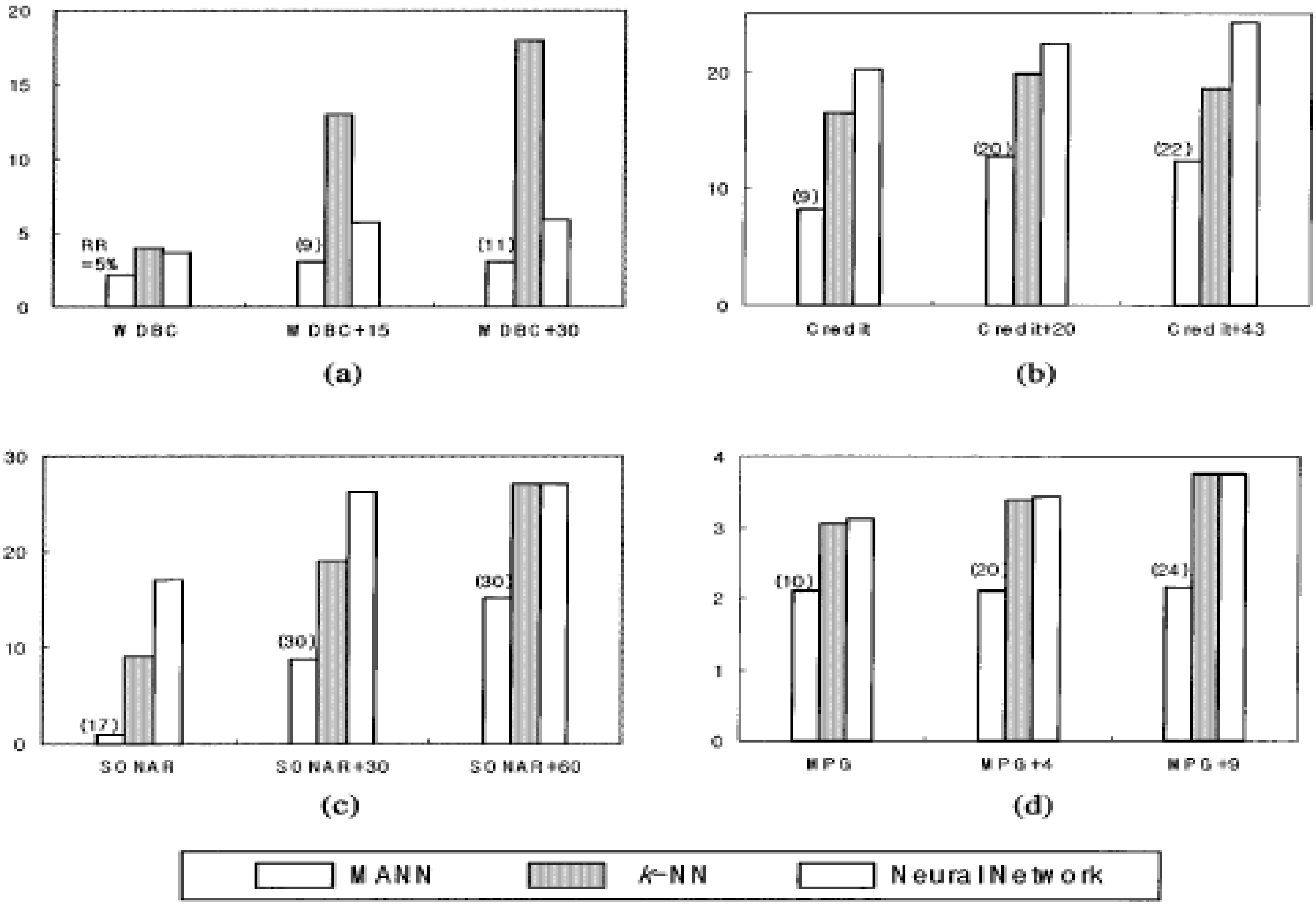
(b)

Experimental results

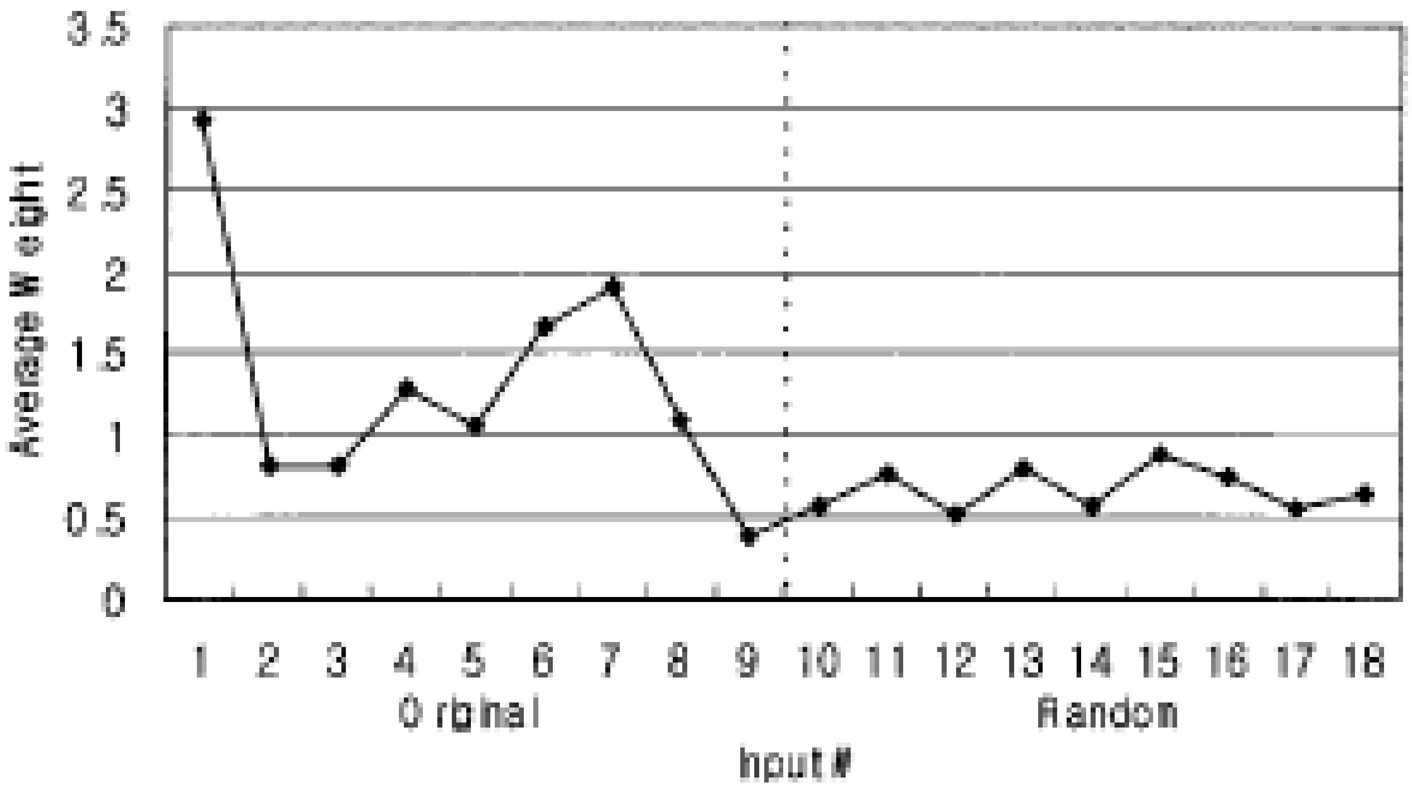
TABLE IV
COMPARISON OF FEATURE WEIGHT LEARNING ALGORITHMS ON PARITY AND SINUSOIDAL PROBLEM

Max. Prediction Accuracy (%)	Parity	Sinusoidal
k-NN	69.2	77.6
Sensitivity	99.1	85.8
Activity	99.2	88.3
Saliency	99.0	87.9
Relevance	99.8	86.0
Relief-F	100.0	80.1
k-NN (VSM)	100.0	88.6
CCF	70.5	68.5
VDM	70.9	68.5
MVDM	71.1	68.4
MI	71.4	73.0

Experimental results



Experimental results



Average feature weights computed in the auto-mpg task

(regression problem)

Experimental results

TABLE V
COMPARISON OF MANN AND OTHER EXPERIMENTS IN REAL WORLD DATASETS

	MANN Error Rate (Rejection Ratio)	Other Experiment
WDBC	2.1% (5.0%)	2.5% (Bennett [31]), MSM-Tree
Credit	8.3% (9.1%)	14.13% (Quinlan [28]), Bagged C4.5
Sonar	1.2% (17.3%)	4.8% (Hastie and Tibshirani [8]), DANN 6.2% (Setiono and Liu [15]), selected features
MPG	2.09 (10.1%)	2.11 (Quinlan [30]), Model Tree

Conclusion

- provides the most similar cases for the prediction query, which demonstrates its explaining capability.
- suggests a hybrid approach to the feature weighting problem with NN technique.
- can be directly applied to classification and regression without additional transformation mechanisms.
- provides on-line learning property.

Limitations

- may not be appropriate to situations that require a short learning time.
- does not explicitly provide any symbolic knowledge.

Future work

- Parameters
  - Length of the case lifetime
  - Treatment of old cases
  - Point to renew the feature weight
- Dynamic behavior
  - Discards old cases
  - Updates the feature weight