Printed in the U.S.A. All rights reserved.

Copyright 2009 © Vocational Evaluation and Career Assessment Professionals (VECAP)

The Neural Network Approach in Vocational Assessment Data Processing

Kanetoshi Hattori
Kobe City College of Nursing

Summary

Neural networking is a data mining method used for clustering, classifying, and estimating. Data processing traditionally carried out by multivariate regression analysis, discriminant analysis, and clustering analysis, has been replaced by neural networking. The traditional methods often presuppose very rigid conditions, such as uni-modal distribution, non-existent outliers within a distribution, and other conditions of normal distribution. Rehabilitation data often has characteristics of an abnormal distribution. Our client population tends to belong to the left end of the general population. At this end, test information is acquired from a very small number of test participants and most of the information is extrapolated from central areas of general population participants. Neural networking remains unused in the rehabilitation field; but it is an effective method to process data for abnormal distributions so common in the rehabilitation field. This presentation proposes use of neural networking in the field of rehabilitation where abnormal distributions are common.

What is an Artificial Neural Network?

Artificial neural networks are computer simulations of electrochemical activities in the biological neural network, a structure that processes signals by means of a mathematical model. The smallest signal processing element in the biological neural network is the neuron. The neural network is composed of multiple neurons connected together. Neural networking has a long history of development and is a product of twists and turns by many researchers.

Artificial neural networks are computer simulations of electrochemical activities in the biological neural network, a structure that processes signals by means of a mathematical model. The smallest signal processing element in the biological neural network is the neuron. The neural network is composed of multiple neurons connected together.

The History of Neural Networking

It is said that the idea of neural networking was first proposed by Alexander Bain (1818 - 1903) in ***Mind and Body: The theories of their relation*** (1873). However, Bain's model was just a diagram—not a mathematical model. The first mathematical model, which used exclusive OR logic, was proposed by N. Rashevsky in his work, ***Mathematical Biophysics*** (1938). Research continued and Warren McCulloch and Walter Pitts proposed a threshold type of neural networking in ***A Logical Calculus of Ideas Immanent in Nervous Activity*** (1943). In 1949, Donald O. Hebb proposed a learning rule for neural connections, known as the Hebb rule. Rosenblatt then made the first electronic neural network (1960) by applying McCulloch & Pitts model and the Hebb learning rule.

Neural networking then faced a stagnant period until contributions from innovative researchers, such as Teuvo Kohonen (1981a, 1981b), Hopfield & Tank (Amari, 1993), Shun'ichi Amari (1978), and others, resulted in an explosion in the field. Neural networking went from the laboratory to the business world. Neural networking has been used in credit ratings, high speed productions, grain inspections, submarine detection, missile guiding, and other areas. However, often the use of neural networking is not announced publicly to prevent the use of this technique by competitors.

The Structure of a Biological Neuron

A biological neuron is composed of three parts: dendrites, a soma, and an axon (Fig.1). Input signals coming out of the dendrites of other neurons flow into the axon. The connection between the axon of a pre-synaptic neuron and the dendrite of a post-synaptic neuron is called a synaptic cleft. In the case of biological neurons, signals are sent as neurotransmitters, which are chemical substances, into synaptic clefts. Once a neurotransmitter is attached to the receptor of a post-synaptic connection, a signal is transmitted to the next neuron.

Synaptic Connections

The strength of the transmitted signal from the axon of a pre-synaptic neuron to the dendrites of a post-synaptic neuron is determined by the density of the neurotransmitter that is discharged and the number of receptors in the dendrites of the post-synaptic neuron (Fig.2). The denser the neurotransmitter at the synaptic cleft, and the larger the number of receptors to which the neuron-transmitter is attached at the post-synaptic neuron, the stronger the transmitted signal from the pre-synaptic neuron to the post-synaptic neuron becomes. Synapses adjust the connection strength between two neurons.

Functions of a Neuron

The signal from a pre-synaptic neuron (Triangles in Fig. 3) is adjusted at the synapse (Hatched area in Fig. 3), then sent to the soma through the dendrites (Rectangle in Fig.3.). The signals sent from the different dendrites are then summed at the soma (Small circles accumulated in basket of Fig. 3.). When the summed signals in the soma exceed a certain level, or the threshold, the soma fires a signal to another neuron through the axon (Stars in Fig. 3.).

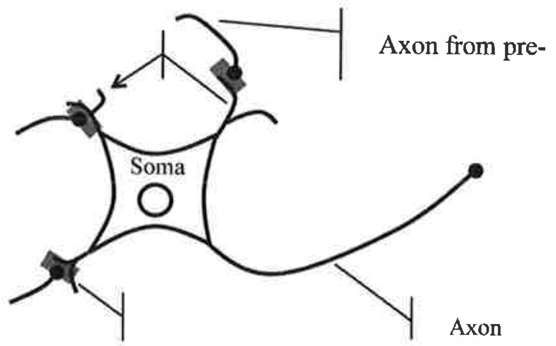

Fig.1 Structure of Biological Neuron

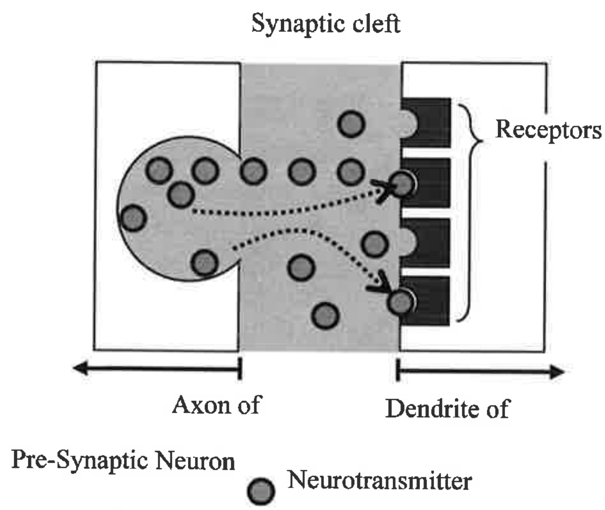

Fig.2

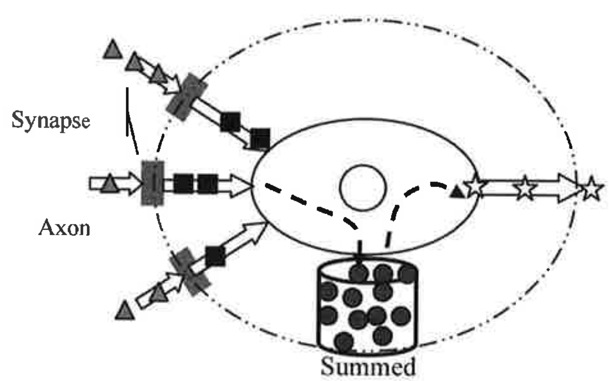

Fig.3 Function of Biological Neuron

Copyright 2009 © Vocational Evaluation and Career Assessment Professionals (VECAP)

The Mathematical Model of a Neuron

Artificial neurons simulate the activities of biological neurons (Usui, Iwata, Kyuuma, & Asakawa, 1995; Baba, Kojima, & Ozawa, 1994; Kamisaka, 1993). (See Figs. 4 & 5.) A signal (I_i) comes from another neuron, which is called the Processing Element (PE), in the artificial neural network. Adjustment at the synapse of the biological neuron is carried out by multiplying signal (I_i) and weight (W_i)--($I_i \times W_i$) (See Fig.4).
Equation: $(\sum(I_i \times W_i)+c)$ This equation represents the summation of a neurotransmitter in a biological neuron. When the quantity of summed signals exceeds the threshold, the processing element fires a new signal (O). Characteristics of output are determined by the types of transfer functions $(f(\sum(I_i \times W_i)+c)$ (See Fig. 5 and section of Transfer Function).

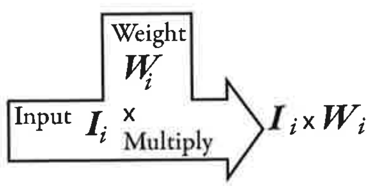

Fig. 4 Connection Adjustment

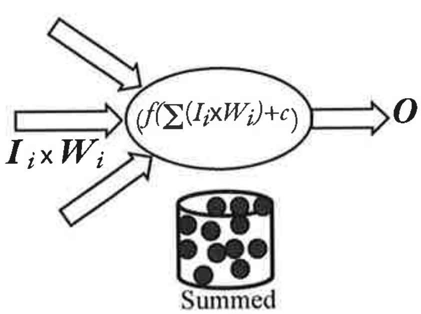

Fig.5 Function of Artificial Neuron (Processing Elements)

Organizing Processing Elements

A biological neural network is made of millions of neurons organized to perform specific functions. Processing elements of an artificial neural network are also organized to carry out computation purposes of the neural network. The organizations are largely classified into two types: one which learns from a supervisory signal and the other which learns from an internal relation of a neural network without a supervisory signal (Acoustic Laboratory, Dept. of Information Environment Integration & Design, Tokyo Denki University, 2006). Fig. 6 shows a neural network with a supervisory signal and the other without a supervisory signal. A typical neural network with a supervisory signal is called backpropagation. Processing elements are connected to elements on a higher layer and the neural network learns weights so that the neural network outcome coincides with the supervisory signal.

A typical neural network that learns from structures of a neural network is the Self Organizing Map (SOM) (Kohonen network). Relationships between processing elements influence SOMs to learn and converge the entire network into a condition of neural connections. These neural connections produce special network characteristics. In other words, the network learns from mutual relations between processing elements.

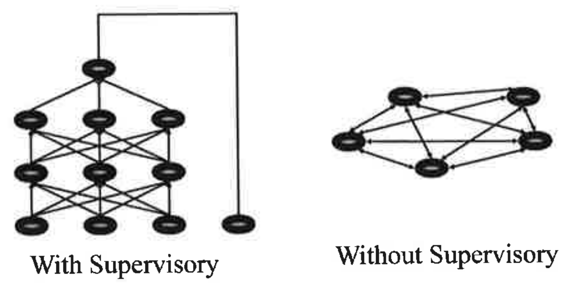

Fig 6. Two Types of Neural Network

The Backpropagation Neural Network

Neurons are interconnected in a biological neural network. Processing elements of artificial neural network are also systematically connected to process data. One of the types of artificial neural networks is backpropagation, which was first described by Paul Werbos, in 1974, and further developed by David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, in 1986 (Wikipedia, 2007).

Backpropagation is a feed forward network that uses supervised learning techniques.

This is an effective method applied to estimation and classification, which were traditionally carried out by regression and discriminant analysis, respectively. Often, performance of backpropagation neural networks is better than traditional methods.

The following (Fig. 7) is a diagram of a backpropagation neural network composed of one input layer, two hidden layers, and an output layer. Numbers of processing elements are 4, 4, 4, and 1 for one input layer, the first hidden layer, the second hidden layer, and the output layer, respectively.

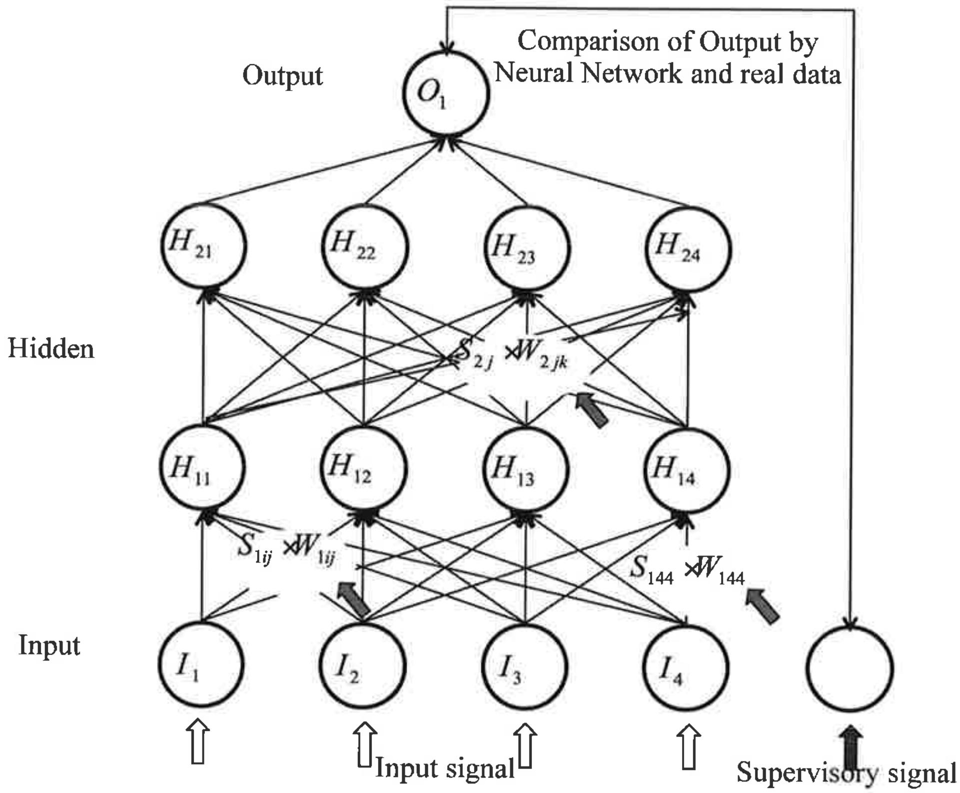

Fig.7 Structure of Backpropagation Data Processing

The Function of Backpropagation Neural Networks

Suppose there are 4 subtest scores, or 4 input signals: Arithmetic Reasoning, Vocabulary, Three-Dimensional Space, and Form Matching. Then, assume that the supervisory signal is the experienced vocational evaluators' judgment of employability, expressed either as 0 (not employable) or 1 (employable). These five signals are entered into the backpropagation neural network. The signal values of the input signals are sent to the first hidden layer.

The 1st input data entered into the left end processing element (I_1) is then sent to the 1st and 4th Processing Elements (H_1, H_{11}, H_{12}, H_{13}, to H_{14}) of the first hidden layer. Each Processing Element of the first hidden layer outputs from 4 Processing Elements. The outputs are multiplied by weights, then summed ($\sum(Ii \times W1ij)$). The read Input value of the *i*th Processing Element on the Input layer is then sent to the *j*th Processing Element of the first Hidden layer. There, the input is multiplied by weight, connecting the *i*th Processing Element of the Input layer and *j*th Processing Element of the first Hidden layer).

Each Processing Element in the first hidden layer outputs signal values according to the equation: $f(\sum(Ii \times W1ij)+c)$ A signal from the input of the Processing Element (H_{1j}) is sent to 4 Processing Elements (H_{2k}) in the second hidden layer. The same data processing cycle is repeated and the Processing Elements (H_{2k}) produce signals for the Processing Element in the Output layer.

Signals from 4 Processing Elements on the second hidden layer are adjusted by weights and summed at the Processing Element on the Output layer (O_1). The

Printed in the U.S.A. All rights reserved.
Copyright 2009 © Vocational Evaluation and Career Assessment Professionals (VECAP)

transfer function of the Processing Element on the Output layer (O_1) produces a value that is an estimation of employability by the Backpropagation neural network. The output from the backpropagation neural network is then compared with actual employability.

If the neural network output does not coincide with the actual value of employability or the value of the supervisory signal, usually, the output signal and the supervisory signal do not coincide with each other at the initial stage of learning. The Backpropagation neural network feed forward then directs weights to change. This learning process is continued until all of the output signals from the neural network coincide with all of the supervisory signals. The display below shows an example of the backpropagation (NeuralWare, Inc., 2000).

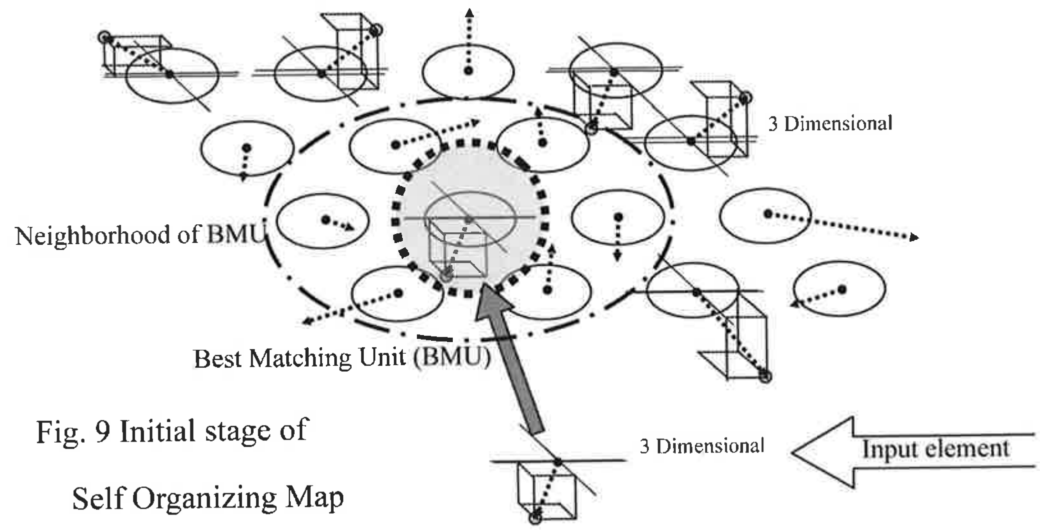

Fig. 9 Initial stage of Self Organizing Map

Backpropagation and Statistical Analyses

Statistical analyses, such as multivariate regression analysis, discriminant analysis, and other techniques, were traditionally used. However, Backpropagation has an advantage over these traditional methods.

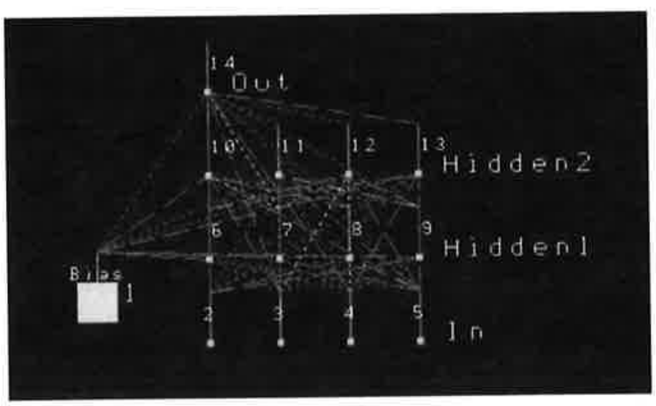

Fig. 8 Display Example of Backpropagation
By Courtesy of NeuralWare Inc.

Backpropagation is often more robust than traditional methods. Estimations by linear regression functions are not as high, though. Estimations can be improved by replacing the linear function with a high-dimensional function. Performance is reversed with cross validation using different samples from the same population. Estimations using high-dimensional functions drop sharply in cross validation; although, the drop in the linear functions is not as large as that of the high-dimensional functions. High-dimensional functions are not as resistant to outliers.

The number of weights used in the backpropagation represented in Fig. 7 is 16+16+4=36. In cases where the multivariate linear regression analysis is applied to the same data set, the number of weights is only 4+1=5. The backpropagation is able to handle many situations, such as outliers by the weights included in the network, which produce flexibility of data processing.

Criteria, such as grades of job performance, are provided through experiences. This criterion tends to be nominal or ordinal data. In cases with ordinal data, the distance between one scale point and another scale point are probabilistically different within the entire scale's span. Such probabilistic characteristics are processed by neural networks.

Hattori, Chan, & Fujita (2001) developed WAIS-R short form by backpropagation. The neural network outperformed traditional linear regression method. Hattori, Yonemoto, & Chan (1991) applied the backpropagation in determining employability of people with disabilities.

Self Organizing Maps (Kohonen Network)

Self Organizing Maps (SOM) are neural networks proposed and practically applied by Teuvo Kohonene in 1981 (Honkela, 1997). The SOM, proposed by Atsumi in 2007, was developed using functions of the visual cortex and autonomously acquires the capability to classify sets of input patterns based on similarities. Learning Vector Quantification, which is often used practically for classification, is a hybrid neural network of the SOM and supervises neural networks.

Each data input set is named and has a 3 dimensional vector (See Table 1.). (If the number of tests which must be processed increases, the number of vectors also increases.)

At the initial stage, an SOM is composed of 16 nodes (4 by 4). A data set with a 3 dimensional vector is input into the SOM. At the beginning of neural network learning, nodes are randomly arranged. Directions of vectors are random, as seen in Fig. 9. The SOM searches for the Best Matching Unit—or, in other words, the SOM finds a vector that is closest to the input vector. As learning progresses, the SOM gathers similar vectors near its neighbor. Finally, the SOM in which different vectors are located at furthest point and similar vectors are located at its neighbor is acquired. The SOM is used as a clustering technique.

Fig. 10 shows an image of the SOM nodes after learning is complete. Vectors were random before learning, but were arranged with a certain nomothetic structure. Input data is organized in a certain cluster.

Table 1. Examples of Vectors Processed

Name	X (Subtest A)	Y (Subtest B)	Z (Subtest C)
John Smith	102	112	89
Mary Cohen	97	103	85
Betty Brown	87	115	120
. . .	. . .	. . .	. . .
Alex Yamada	67	95	122

Similar vectors are agglutinated around the neighborhood and different vectors are placed at distant locations. (The 3 dimensional vectors in Fig. 9 are simplified as 2 dimensional vectors for easier understanding.)

Learning Vector Quantification (LVQ), which is a hybrid of the SOM and the Backpropagation, is often used for data classification. Cluster names are given as supervised signals and data input is processed by the SOM. Fig. 11 is a display example of a LVQ (NeuralWare, Inc., 2000).

SOMs and Hierarchical Clustering

The hierarchical clustering approach has been previously used; but this method has a principle problem. The hierarchical clustering constructs in incremental steps. Data, which is already included in larger clusters, will not compose another cluster with other data included in the whole data set. In case of *k*-means or SOMs, it is possible to reorganize whole clusters.

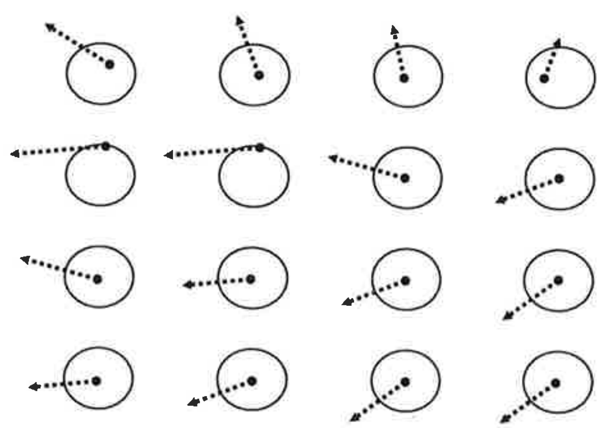

Fig. 10 SOM after Learning

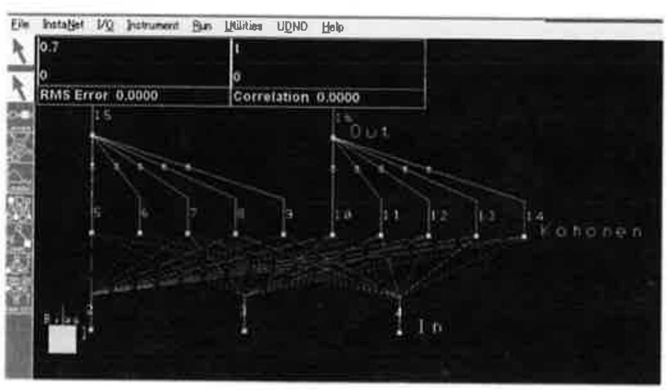

Fig. 11 Learning Vector Quantification
By Courtesy of NeuralWare Inc.

Transfer Functions

There are advantages to using transfer functions in neural networking. A transfer function is a function that determines the characteristics of an output signal (NeuralWare, 2000). The neural network, at its early stage, started from *Threshold Step transfer function,* which is appropriate for an 'all or none' principle (0 or 1). The *Piecewise Linear function* is a function $f(x)$ whose definition is given differently on disjoint subsets of its domain. This method is appropriate for outputs that have different noncontiguous characteristics. The *Sigmoid function* is defined as $f(x)=1/(1+e^{-x})$ and its asymptotic lines are $y=0$ and $y=1$. The *Hyperbolic tangent function* is defined as $f(x)=(e^{x}-e^{-x})/(e^{x}+e^{-x})$ and its asymmetric lines are $y=-1$ and $y=1$. Lastly, the *Gaussian function* is defined as $f(x)=a^{-(x-b)^2/c^2}$ for some real constants a>0, b, and c. The neural network is able to select network output distribution which is similar to actual distribution.

Data Distribution in Rehabilitation

It is often observed that rehabilitation data does not have the characteristics of a normal distribution. Outliers are observed in the distribution and the data is often skewed. There are cases where the rehabilitation data does not pass the Kolmogorov-Smirnov test.

It should be acknowledged that data from rehabilitation clients often belongs to the left end of the norm produced by the general population. The actual numbers of samples used in such a population is very small. In fact, the amount of information in these samples is less than 5% of the entire sample population (Carnap, 1949, 1958; BarHillel, 1964). Most of the information used in this population is from extraporation of general population.

Characteristics of rehabilitation clients vary. It is common practice in test development to sample data based on the normal distribution. It is advised to add profile data to the general norm data.

Steps in Neural Network Analysis

A neural network is a complex system and it is difficult to track the process flow because there are so many weights used in data processing. Therefore, it is almost impossible to interpret the relationship between variables, weights, and outcomes. The following method is often used. See Fig. 13.

1. Randomize sample data from a certain population.
2. Separate the randomized data into two data sets: one used for learning and the

Copyright 2009 © Vocational Evaluation and Career Assessment Professionals (VECAP)

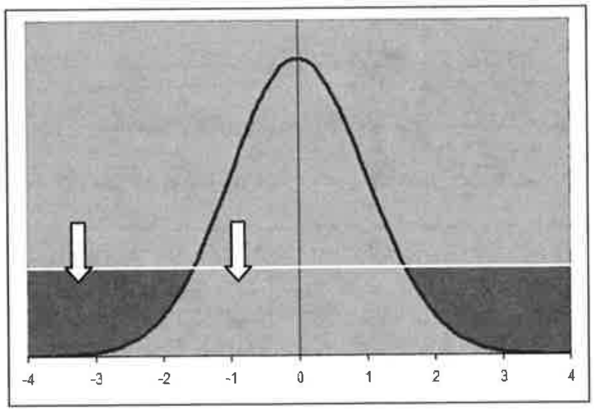

Fig. 12 Normal Distribution and Distribution areas to which information should be added.

other used as the performance test of the neural network outcome. Custoarily, the percentage is 50% fo learning and 50% for testing, or 70% for learning and 30% for testing.

3. A neural network model is constructed The number of processing elements, the number of hidden layers, the type of transfer function (Often, the default setting will perform well.), and the number of learning cycles are factors to be determined.
4. Neural network learning is performed only by the learning data set and a set of weights. Once correlations are complete, the error rates are saved (the learning, or validation, phase).
5. The data set to be tested is processed by the weights saved in the learning phase. (Test, or cross-validation, phase)
6. If the result is not acceptable, one must return to the construction of the neural network model. However, if the result is acceptable, one returns to randomize and re-analyze the data using the same model.

Once the analysis from the neural network is complete, similar data sets are often analyzed using traditional statistical analyses to evaluate the advantages of the neural network. Hattori, Chan, & Fujita (2001) compared results of backpropagation neural networks with regression analyses and reported superiority of the backpropagation.

Risks in Using Neural Network

There is a risk of over-learning in neural networking. Since neural networks use large numbers of weights to process data, excessively high performance is acquired. When the number of data sets is not large enough, performance drops sharply at test phase, or cross validation.

Hattori, K., Kosaka, M., Iwqmoto, S., Washida, M. & Ikeda, S. (2006) estimated incidences of Cardio Vascular Accidents from the age cohort data in 3,368 cities and townships acquired in the Japanese census. In order to improve estimation performance, the 3,368 locations were initially classified into 15 clusters by the *k*-means method. Table 2 shows the Cluster number, N of locations, the number of learning cycles reaching outcomes by neural networks, and correlations between the estimated number of CVA incidences by a neural network and the actual statistical data. When N is too small, the r is excessively high and over-learning occurs. The neural network is not a versatile machine; but it is a powerful tool.

Data Processing in the Past

Statistical methods that demand rigid hypotheses for normal distributions have been used in rehabilitation. It was not a common practice to carry out analyses of normality, such as the Kolmogorov-Smirnov test. Exploratory data analysis, a simple way to find skewness, multi-modality, and outliers, was not used. Multivariate analyses are fragile against outliers, due to their mathematical structures, and were used without considerations to compensate their weaknesses. In the past, there were no alternatives.

Dramatic Changes for the Future

Although clear figures are not yet known, it is commonly understood among professionals that there will be drastic changes in medicine within the next 5 to 10 years. Progress in brain science and omics are predicted to be key. The outcomes of these sciences, and the technologies developed to carry out the sciences, are changing the methodologies in rehabilitation related fields.

Rehabilitation professionals have used personality tests based on a behavioralism

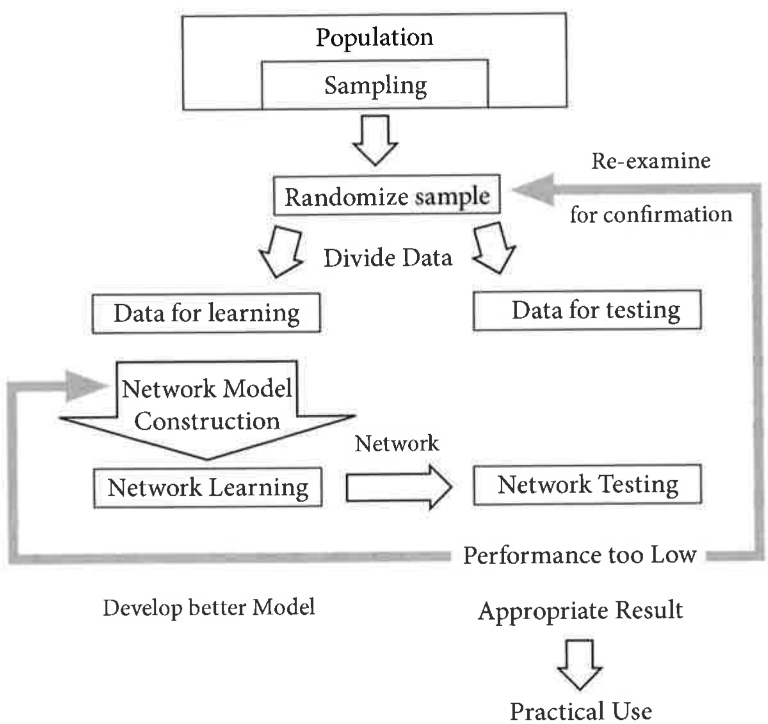

Fig. 13 Procedures in Neural Network

Table 2. Correlations between Actual Death Rates by CVA and their Estimation by Neural Network

Cluster #	N	Number of Learning Cycle	r
1	295	300 thousand	0.922
2	456	500	0.903
3	397	500	0.908
4	95	200	0.999 *
5	291	300	0.944
6	395	600	0.913
7	339	300	0.918
8	235	200	0.962
9	354	800	0.904
10	41	50	0.999 *
11	200	400	0.932
12	5	Unable to converge	
13	19	50	1.000 *
14	17	50	1.000 *
15	208	300 thousand	0.956

* Possible over-learning

paradigm. A paradigm shift is observed in personality testing and measurement technology. Cloninger (1993; 1994a; 1994b; 1996) proposed personality testing based on research on genomes and neurotransmitters. Cloninger's model may not be the major model in vocational assessment; yet, the model has been used in psychiatric treatments. Recent clinical assessment is conducted based on levels of serotonin, dopamine, and noradrenaline (Shirakawa, 2004). There has also been research relating behavioral parameters, such as disconcertedness, anxiety, depression, etc. with the involvement of monoamines (Kasahara, 1998). It is only a matter of time before genome information is matched with behavioral data.

Numerous technologies regarding genome analysis have been developed. Some methods, such as a method to display two kinds of clustering analyses in a two-dimensional array, can be applied immediately in rehabilitation. Dr. Hattori is using genome ontology (Takai, 2006) as a technique in textmining nursing data. Even a very primitive application of a three-dimensional rendering technique (Kitaoka & Iwata, 2006) to statistical data surprises rehabilitation professionals. The three-dimensional graphic display is a very effective method in data exploration.

Tests developed in cognitive psychology that are specific to certain locations of the brain are based on clinical case studies and imaging technology. The Paired Associates Learning subtest of CANTAB targets the parietal and temporal lobes; thus, it is useful in detecting early stages Alzheimer's (Swainson R. et al., 2001). Data distributes such as these are abnormal.

In order to pursue new fields, we need a new, symbolic system to represent objects. Newtonian physics would not have been constructed without calculus as the symbolic system to represent the force of gravity. To make more progress, we need another symbolic system to represent objects. In ***Tractatus Logico Philosophicus***, Ludwig Wittgenstein (1961) said, *the limits of my language* mean the limits of my world. 5.6 Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt.

Printed in the U.S.A. All rights reserved.
Copyright 2009 © Vocational Evaluation and Career Assessment Professionals (VECAP)

References

Acoustic Laboratory, Dept. of Information Environment Integration & Design, Tokyo Denki University (2006) ***Neural Network*** retrieved at 2007/2/19 from http://www.sound.sie.dendai.ac.jp/dsp/Text/PDF/Chap5-3.pdf. (Text in Japanese)

Amari, S. (1978). ***Mathematics of Neural Network.*** Saogyo tosho, Tokyo. (Text in Japanese)

Atsumi, M. (2006). Self Organizing Map retrieved 2007/2/19 from http://www.intlab.soka.ac.jp/~matsumi/material/NC/SOM/SOM.html.

Baba, N., Kojima, F., & Ozawa, S. (1994) ***Foundation and Application of Neural Network***, Kyoritsu-Shuppan, Tokyo. (Text in Japanese)

Bain, A. (1873) ***Mind and Body. The Theories of Their Relation*** cited in Olmsted, D.D. (2006) ***History and Principles of Neural Networks to 1960*** retrieved 2007/2/19 from http://www.neurocomputing.org/NNHistoryTo1960.aspx.

Bar-Hillel, Y. (1964). ***Language and Information***, Reading, MA., Addison-Wesley Publishing Co.

Carnap, R. (1949). ***Meaning and Necessity***, Chicago, University of Chicago Press.

Carnap, R. (1958). ***Introduction to Symbolic Logic and its Applications***, New York, Dover Publications.

Cloninger, C.R., Svrakic, D.M., & Przybeck, T.R. (1993) A psychobiological model of temperament and character, ***Archives of General Psychiatry***, 50, 975-990.

Cloninger, C.R. (1994a) Temperament and personality. ***Current Opinion inNeurobiology*** 4, 266-273.

Cloninger, C.R. (1994b) The genetic structure of personality and learning: a phylogenetic model. ***Clinical Genetics***, 46, 124-137.

Cloninger, C.R., Adolfsson, R., & Svrakic, D.M. (1996) Mapping genes for human personality, ***Nature Genetics*** 12, 3-4.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. Personality and Social Psychology Bulletin, 31, 166–180.

Hattori, K., Chan, F. and Fujita, K. (2001) Backpropagation neural network short form of WAIS-R Japanese edition, ***International Journal of Rehabilitation Research,*** 24(4), 329-335.

Hattori, K., Kosaka, M., Iwamoto, S., Washida, M. & Ikeda, S. (2006) Death rate of CVA from age cohort data of Japanese census data by k-means clustering and backpropagation, Poster presentation at ***2006 National Assembly of Japanese Academy of Nursing Science.***

Hattori, K., Yonemoto, K., & Chan, F. (1992) An application of Neural Network in Vocational Evaluation, A Development of toy Model, ***Vocational Rehabilitation***, 5, 66-70. (Text in Japanese)

Hebb, D.O. (1949). ***The Organization of Behavior***, John Wiley & Sons, New York cited in Olmsted, D.D. (2006) ***History and Principles of Neural Networks to 1960*** retrieved 2007/2/19 from http://www.neurocomputing.org/NNHistoryTo1960.aspx.

Hino, Washida, Ikeda, & Hattori (Under-review) A Development of Foot measurement method using Neural Network (3D estimation model) for appropriate nurse shoe selection for nurse students. Poster presentation at ***Annual assembly of Japan Academy of Nursing Education.***

Honkela, T. (1997) ***Self-Organizing Maps in Natural Language Processing*** retrieved at 2007/2/19 from http://www.cis.hut.fi/~tho/thesis/index.html.

Hopfield, J.J. & Tank, D.W. (1985) Neural Computation of Decisions in Optimization Problems, ***Biological Cybernetics***, 52, 141-152 cited in Amari, S. (Ed.) (1993) ***Information & Computing – 75, New Expansion of Neural Network***, Science, Tokyo (Text in Japanese)

Kamisaka. Y. (1993) ***Mathematical Foundation of Neuro Computing***, Kindai-Kagakusha, Tokyo. (Text in Japanese)

Kitaoka, H. & Iwata, Y. (2006). Modeling of internal organ and dynamic simulation, ***Clinical Genome Informatics Manpower Development Unit, Special Course of Genome Informatics Training Manual***, 2(14), 1-30., Foundation for Biomedical Research and Innovation, Kobe City

Kohonen, T. (1981a). Construction of similarity diagrams for phonemes by a self organizing algorithm. ***Technical Report TKK-F-A463,*** Helsinki University of Technology, Espoo, Finland cited in Honkela, T. (1997) ***Self-Organizing Maps in Natural Language Processing*** retrieved at 2007/2/19 from http://www.cis.hut.fi/~tho/thesis/index.html.

Kohonen, T. (1981b). Hierarchical ordering of vectorial data in a self-organizing process. ***Report TKK-F-A461***, Helsinki University of Technology, Espoo, Finland cited in Honkela, T. (1997) Self-***Organizing Maps in Natural Language Processing*** retrieved at 2007/2/19 from http://www.cis.hut.fi/~tho/thesis/index.html.

McCulloch, W.S. & Pitts, W.H. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity, ***Bulletin of Mathematical Biophysics***, 5:115-137 cited in Olmsted, D.D. (2006) ***History and Principles of Neural Networks to 1960*** retrieved 2007/2/19 from http://www.neurocomputing.org/NNHistoryTo1960.aspx.
NeuralWare, Inc. (2000) NeuralWorks, ***Professional II/Plus***, Author, Carnegie, PA.

Rashevsky, N. (1938) ***Mathematical Biophysics*** cited in Olmsted, D.D. (2006) ***History and Principles of Neural Networks to 1960*** retrieved 2007/2/19 from http://www.neurocomputing.org/NNHistoryTo1960.aspx.

Rosenblatt, F (1960). Perceptron Simulation Experiments, ***Proceedings of the IRE, 48:301-309*** cited in Olmsted, D.D. (2006) History ***and Principles of Neural Networks to 1960*** retrieved 2007/2/19 from http://www.neurocomputing.org/NNHistoryTo1960.aspx.

Swainson R., Hodges J.R., Paykel E.S., Galton C.J., Semple J., Michael A., Dunn B.D., Iddon J.L., Robbins T.W., Sahakian B.J. (2001). Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. ***Dementia and Geriatric Cognitive Disorders*** 12: 265-280.

Shirakawa, O. (2004). Current practice of drug treatment, ***Clinical Neuroscience***, 22(2), 202-207.

Takai, T. (2006). An Introduction to Genome Ontology, ***Clinical Genome Informatics Manpower Development Unit, Special Course of Genome Informatics Training Manual***, 1(13), 1-48. , Foundation for Biomedical Research and Innovation, Kobe City.

Usui, S., Iwata, A., Kyuuma, K. & Asakawa, K. (Eds.) (1995) ***Foundations and Practices, Neural Network***, Korona-sha, Tokyo. (Text in Japanese)

Wittgenstein, L. (1961) ***Tractatus Logico Philosophicus***, Routledge & Kegan Paul, Ltd. **A Substance Use Screening Tool for Rehabilitation Counselors Working in Vocational Rehabilitation Agencies**