Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 1–11
Expert Systems with Applications
www.elsevier.com/locate/eswa

Artificial neural networks classification and clustering of methodologies and applications – literature analysis from 1995 to 2005

Shu-Hsien Liao [a], Chih-Hao Wen [b]

[a] Department of Management Sciences and Decision Making, Tamkang University, No. 151, Yingjuan Road, Danshuei Jen, Taipei 251, Taiwan, ROC
[b] Graduate School of Resource Management, National Defence University, Management College, P.O. Box 90046-17 Jon-Ho, Taipei County, Taiwan, ROC

Abstract

Based on a scope of 10,120 articles on ANNs, this paper uses data mining including association rules and cluster analysis, to survey these ANNs papers through keyword classification and clustering of articles from 1995 to 2005, exploring the ANNs methodologies and application developments during that period. The four decision variables of keywords, author's nationality, research category, and year of publication, are implemented for data mining with total of 110,080 data items. The research findings show that some specific ANNs methodologies and applications pattern can be extracted from the mining results, and these describe the ANNs development over this period. In addition, using more data mining approaches for analysis could provide different explanations for ANNs development. Finally, discussion and brief conclusion are presented.

Keywords: Artificial neural networks; Artificial neural networks methodologies and applications; Data mining; Association rule; Cluster analysis; Literature analysis

1. Introduction

Artificial neural networks (ANNs) are computational modeling tools that have been extensively used in many disciplines to model complex real-world problems. ANNs may be defined as structures comprised of densely interconnected adaptive simple processing elements (called artificial neurons or nodes) that are capable of performing massive parallel computations for data processing and knowledge representation (Schalkoff, 1997). Although ANNs are drastic abstractions of their biological counterparts, the idea of ANNs is not to replicate the operation of the biological systems but to use what is known about the functionality of the biological networks in order to solve complex problems. Neural networks discover patterns and relationships in massive amounts of data by using hardware and software that simulates the processing patterns of the human brain. Contrary to conventional computing methods, ANNs are 'trained' to produce the desired input–output relationships. During the training (learning) phase, examples of data are presented to the network and, using a learning algorithm, the parameters are tuned to adjust the network behavior. According to available knowledge of the problem, and the objective of the operator, the learning procedure employed can be either 'supervised', 'unsupervised', or both. The supervised learning procedure is performed with pairs of known input–output patterns, whereas unsupervised learning consists of presenting training examples to the network input while the network organizes itself progressively to reach maximal separation between the naturally occurring classes of examples. The principal applications of ANNs have been in the area of pattern recognition. Here, a pattern is turned into a

E-mail addresses: michael@mail.tku.edu.tw (S.-H. Liao), 6060101zq-az@yahoo.com.tw (C.-H. Wen).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.11.014

feature vector, which is then used as ANN input, and the output is interpreted as identifying the input to be a member of one of a number of classes of possible inputs. An important quality of neural networks (referred to as generalization) is that when they are correctly trained, neural networks can appropriately process data that have not been used for training. The most popular neural networks are multilayer perceptions, which are generally supervised–trained with the error back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986). One major property of these networks is their ability to find non-linear surfaces separating the underlying patterns, which is generally considered to be an improvement on conventional methods.

On the other hand, radial basis function (RBF) network is a particular class of multilayer networks (Poggio & Girosi, 1989), where learning occurs usually in two stages: learning in the hidden layer (usually by an unsupervised bottom–up self-organizing method such as *K*-means clustering) followed by the output layer (a top–down supervised method such as least squares estimation). RBF networks have two important advantages: finding the input to output map using local approximators, and rapid learning requiring fewer examples. Another popular class of networks is the self-organizing map, or Kohonen network (Kohonen, 1988). A Kohonen network consists of two fully connected-unit layers. The output layer is generally ordered in a low-dimensional framework (a one-dimensional array or a two-dimensional matrix) of units. The objective of this network is to build a map where units of an area are activated when inputs with similar characteristics are presented. Among the other popular networks are adaptive resonance theory (ART) networks and their derivates (ART1, ART2, fuzzy ART, etc.) (Carpenter & Grossberg, 1992) and Hopfield models (Hopfield, 1982). One quality of neural networks is that they can be considered as non-linear statistical methods. Nevertheless, a large amount of data is required to overcome the existing non-linearities in the data structure. Therefore, the attractiveness of ANNs comes from the remarkable information processing characteristics of a biological system such as non-linearity, high parallelism, robustness, fault and failure tolerance, learning, ability to handle imprecise and fuzzy information, and their capability to generalize.

This research began on June 2005, and it was first based on a search in the keyword index of article for 'neural networks' on the Elsevier SDOS, IEEE Xplore, Blackwell Synergy, Wiley InterScience, and Taylor & Francis online database, for the period from 1995 to 2005, in which 10,120 articles were found on 30 June 2005. Based on a scope of 10,120 articles on ANNs, this paper uses a data mining approach including association rules and clustering analysis to survey the ANNs through keyword classification and clustering of articles from 1995 to 2005. Based on this, the ANNs methodologies and applications development during that period are explored. The rest of the paper is organized as follows. Section 2 introduces research design. Sections 3 and 4 describe the data mining process and its results. Section 5 discusses our findings, and Section 6 contains a brief conclusion.

2. Research design

2.1. Why keywords are used as data source for literature analysis?

Basically, there are three aspects for research, including research problem, theory, and methods/tools. Sometimes, the research objective and motivation is described as the research problem, which contains the nature of a specific problem domain. In addition, a problem domain might be a person, a group of people, an organization, an object, or an event, which can be estimated and evaluated. This means that a problem domain has specific domain knowledge with certain complexity and that this knowledge may exist either tacitly or explicitly under different situations and conditions. This form of domain knowledge could be a threshold to researchers who have limited understanding of this problem domain. This means that researchers should understand or learn something on research problem domain knowledge before starting their research. On the other hand, a research problem may describe a process or procedure of research experiment, case study, system development, or modeling on a specific research field. These processes or procedures focus on a limited scope that social science or natural science can manipulate and observe the research problem by collecting and analyzing data from the problem domain. In this regard, a study must define the research process or procedure in order to illustrate the research scope from real-world situations or conditions. A research problem is a logical description to show why a study is critical or what is its contribution and originality in a specific problem domain. Without good enough domain knowledge, the research problem can be the most difficulty part of work at the initial stage of research. Because a researcher must spend great amount of time to understand the domain knowledge deeply and then can provide a basis to understand and define the research scope.

On the other hand, once the scope of a research problem domain has been determined, researcher can describe research problem by citing specific theory. A theory is generalized phenomena, which presents results or findings from specific research experiment or long term observation with systematic approach. However, a phenomenon can be extended to different theories on different research problem domain. For example, in physics, the principle of inertia states that objects continue in a state of rest or of uniform motion unless acted upon by forces. In knowledge management, the knowledge inertia describes that people may stem from the use of routine problem solving procedures, stagnant knowledge sources, and following past experience or knowledge (Liao, 2002). Due to diversity of research problem domains, researchers must to cite theory or read literature in relation to the research problem in order to fit how people refer to a phenomenon in a specific problem domain.

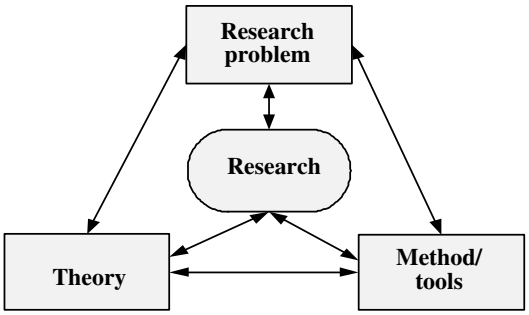

Fig. 1. Aspects of a research.

This might be a problem for a neophyte or in a new research area for look for the appropriate literature in order to reference necessary articles to describe a research problem from the huge mass of journals, magazines, and other research materials. In this regard, a study might be different from others with different theories, even though have similar research phenomenon. Therefore, once a research problem and the chosen theory are combined, the axis of a study has been established. After that, when domain knowledge and theories have been determined to describe a specific problem, certain research methods or tools, such as software, hardware, measurement tools, can be implemented to solve the research problem. At this stage, specific data format collecting from the research problem domain is confined to research methods or tools. Therefore, a researcher must realize the fitness between the data format of a research problem and its methods/tools, and then consider the possibility of explaining the research problem from the experiment results.

In Fig. 1, this paper presents the research design. We consider that a research article contains three aspects including the research problem, theory, and methods/tools. Furthermore, the keywords of a research article usually include these three aspects, which present what problem domain knowledge this research has, how it describes the research problem, and where results come from specific process or procedure with certain methods/tools. Therefore, this might be an approach to reference literature using a keyword index search from an online digital database to help a researcher establish the research framework from the three aspects as soon as possible. Thus, this paper uses article keywords as a data source in order to investigate ANNs methodologies and applications development over the period from 1995 to 2005.

2.2. How the research starts?

On 30 June 2005, this research retrieved 10,474 articles with using the keyword index search for 'neural networks' from the Elsevier SDOS, IEEE Xplore, Blackwell Synergy, Wiley InterScience, and Taylor & Francis online databases, for the period from 1995 to 2005. After that, we categorized four data groups for analysis, including keywords (two categories – methodologies and applications), year for publication (two data groups – before 2000 and after 2000), research category (two categories – science and social science), and first author's nationality. By doing so, this paper keys in those data on Microsoft Excel with designed data classification and categories. During the data processing stage, 354 articles were found to have no keywords and were excluded from processing. Therefore, 10,120 articles were considered in this research. Furthermore, the authors of those articles used different keywords to denote similar methodologies and applications, such as 'fuzzy neural network' and 'fuzzy-neural network', or 'forecast' and 'forecasting'. For consistency, this research generalizes all keywords by selecting the most frequency terminology for a keyword when indexing. Finally, total of 110,080 data items were keyed into the database. This research established relational tables on MS Access 2002 and transferred them on the MS SQL Server within an OBDC environment in order to implement data tables on SPSS Clementine.

2.3. Analysis approach – data miming

2.3.1. Association rules and the Apriori algorithm

Association rules have been widely employed in different disciplines to explore relationships between data in the database. If $L = \{l1, l2, \ldots, ln\}$ denotes the set of ANNs methodologies and applications in the database, and both X and Y represent decision variables which are subsets of L and are independent, then an example of an association rule is an implication of the form: $X \rightarrow Y$, where $X \subset L$, $Y \subset L$ and $X \cap Y = \varnothing$. Each association rule has the two measurement standards of support and confidence (Wang, Chuang, Hsu, & Keh, 2004). The support of a rule, denoted by Sup(X), is the percentage of transactions in the database, which contains the itemset $X \cup Y$. All itemsets which satisfy the support threshold, called minimum support, are called large itemsets. During the core phase of the algorithm, all large itemsets are generated. The confidence of a rule, denoted by Conf ($X \rightarrow Y$), is the percentage of transactions, which contain X and also contains Y (Agrawal & Srikant, 1994; Tsechansky, Pliskin, Rabinowitz, & Porath, 1999). In practical investigations, these rules are usually regarded as "interesting" only if the support and confidence exceed certain threshold values (Coenen, Goulbourne, & Leng, 2004). Moreover, the judgment standard is called the lift, which is defined as: Lift = Confidence($X \rightarrow Y$)/Support(Y) (Wang et al., 2004).

The best-known strategy for association rule mining is called Apriori (Agrawal & Srikant, 1994; Kouris, Makris, & Tsakalidis, 2005). Apriori algorithms read the recorded data in the database repeatedly and generate large itemsets after each reading. In order to reduce the calculation load and enhance efficiency, only the support for candidate itemsets are calculated.

2.3.2. Cluster analysis and K-means algorithm

The process of partitioning a large set of patterns into disjoint and homogeneous clusters is fundamental in knowledge acquisition. It is called *Clustering* in the

literature and it is applied in various fields, including data mining, statistical data analysis, compression and vector quantization. The *K-means* is a very popular algorithm and is one of the best for implementing the clustering process. *K*-means clustering proceeds in the following order. Firstly, K number of observations is randomly selected from all N number of observations, according to the number of clusters, and these become centers of the initial clusters. Secondly, for each of the remaining $N-K$ observations, find the nearest cluster in terms of the Euclidean distance with respect to $xi=(xi1,xi2,\ldots;xip,\ldots,xiP)$ is found. After each observation is assigned the nearest cluster, the center of the cluster is re-computed. Lastly, after the allocation of all observation, the Euclidean distance between each observation and cluster's center point is calculated to confirm whether or not it is allocated to the nearest cluster. In addition, the *K*-means algorithm implementing for cluster analysis as data mining approach has been discussed on several researches (Kuo, Liao, & Tu, 2005; Vrahatis, Boutsinas, Alevizos, & Pavlides, 2002).

2.4. Data mining process

This paper proposes the association rule and clustering analysis for data mining to extract knowledge from decision variables including keywords, year of publication, research category, and first author's nationality. Knowledge extracted from these mining results can serve as useful knowledge to survey the development of artificial neural networks over the last decade. Fig. 2 presents the data mining process.

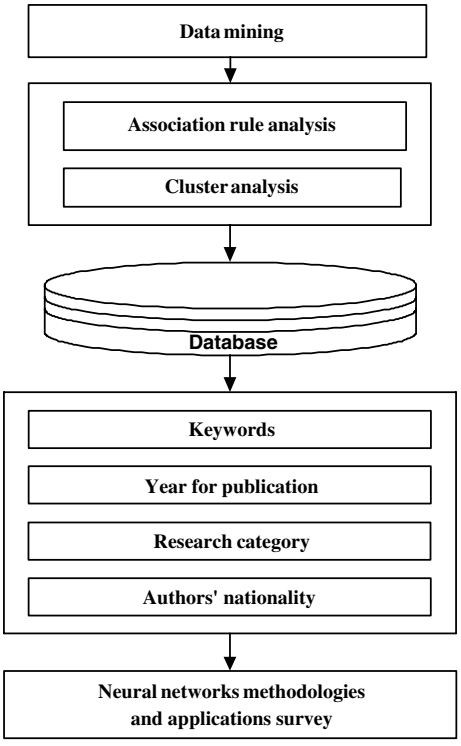

Fig. 2. Data mining process.

2.5. Data mining tool – SPSS Clementine

In this research, SPSS Clementine is employed as data mining tool for association rule and cluster analysis. The difference between Clementine and other software is that its data processing is through the use of nodes, which are then connected together to form a stream frame. In addition, data visualization can present to users after mining process has been done. All of the nodes can be divided into six categories: the source node, record options node, field options node, graphs node, modeling node, and output node (shown in Fig. 3).

Source node: The source selection not only includes those that can undergo the data connection node via the open database connectivity (ODBC) and relational database management system. It also includes nodes that can input all sorts of common file contents.

Record options node: This is for the recording and correction of data. These operations are very important in the data investigation stage of data interpretation and data preparation stage, because these operations enable the data to fulfill specific business needs.

Field options node: The field option node can help the user do modeling and data preparation for the logical data design stage.

Graphs node: This is one of the stages in the data mining process that uses graphs for exploratory data analysis. Another purpose is to examine the new record option's distribution and relationships.

Modeling node: Modeling is the core of the data investigation process. The modeling method of this node enables the user to retrieve new information from the data and to form a forecast model. These modeling methods are derived machine learning (ML), artificial intelligence (AI), and statistics, etc. All these methods have their own advantages and are suitable for specific types of problem. The algorithm includes:

(1) Decision Tree (C5.0, CART).
(2) Neural Net and RBF Function.
(3) Association Rule (Apriori, GRI).
(4) Sequence Detection.
(5) Cluster Analysis (*K*-means, Two-step and Kohonen).
(6) Regression (Linear Regression, Logistic Regression).
(7) Factor Analysis and PCA.

Output node: This offers one method to achieve data that is related to the users and the model. It can output all types of data in different forms to other software interfaces.

3. Association rule analysis

3.1. Keywords and methodologies/applications association rules

This research categorizes keywords on 149 kinds of methodologies and 2359 kinds of applications according

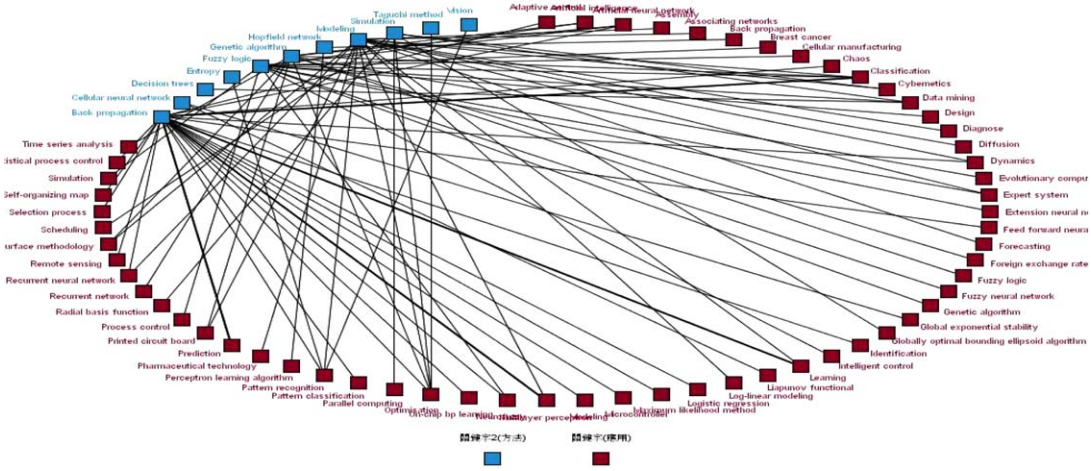

Fig. 3. Web diagram of keywords and methodologies/applications association.

to the classification of research problem (applications), theory (methodologies), and methods/tools (methodologies). Therefore, both methodologies and applications contain diverse classification content. Fig. 3 shows that association rules exist in which two decision variables have more than one-time association.

Technically, by reducing the threshold value of support and confidence, the Apriori algorithm of Clementine can increase the levels of decision variables. However, because there are so many keywords for methodologies and applications, even setting the threshold value to one (antecedent * 1 and consequent * 1), a system can only mine one level of consequent and antecedent. Therefore, this research mines association rules for keywords of methodologies and applications by setting min sup. = 5 and min conf. = 5. By doing so, this research sets 'methodology' as antecedent of association rule and 'application' as the consequent. Thus, the mining results show five association rules of neural networks methodologies and applications in Table 1.

In Table 1, 'Back propagation' and 'Modeling' are two methodologies that show association rules. The methodology 'Back propagation' and the application 'Classification' have a high association result, with the lift value of 2.14. In addition, the methodology 'Modeling' together with the applications 'Associating networks', 'Pharmaceutical technology', 'Response surface methodology', and 'Classification' have highly associated results. In addition, the keyword 'Neural nets' integrated with 'Modeling' and 'Associating networks' and 'Pharmaceutical technology' has a lift value of 17.15. This means that these methodologies and applications are well developed in the ANNs research field.

3.2. Keywords and author's nationality association rules

This paper uses the first author's nationality as a decision variable to investigate the relationship between methodologies/applications and the affiliation location of the author's. Sixty-five countries or areas are used to categorize the first author's affiliation. This means that authors from those countries or areas have been participated in ANNs research, and it shows that ANNs methodologies and applications are main research trend world wide. In addition, this research sets 'Keywords' as the antecedent for association rule and 'Author's nationality' is the consequent, with a threshold value of min sup. = 5 and min conf. = 5. Fig. 4 shows the results of these association rules.

There were 24 rules extracted from the mining results. A total of 12 countries have highly association with ANNs methodologies and applications, including China, Japan, Taiwan, UK, USA, India, Turkey, Switzerland, Brazil, Turkey, France, and Spain. For example, 'China', has association with methodologies of 'Hopfield network', and 'Back propagation', where 'Global exponential stability' is the antecedent of application. Similarly, 'USA' is associated with the methodology of 'Fuzzy logic' and 'UK' is associated with the methodology of 'modeling'. In addition, 'Back

Table 1
Association rules for keywords and methodologies/applications (min sup. = 5, min conf. = 5)

	Support	Confidence	Lift	Consequent	Antecedent
1	7.60	9.10	2.14	Keywords (application) = Classification	Keywords (methodology) = Back propagation
2	5.80	5.10	17.15	Keywords (application) = Associating networks	Keywords (methodology) = Modeling
3	5.80	5.10	17.15	Keywords (application) = Pharmaceutical technology	Keywords (methodology) = Modeling
4	5.80	5.10	10.29	Keywords (application) = Response surface methodology	Keywords (methodology) = Modeling
5	5.80	5.10	1.20	Keywords (application) = Classification	Keywords (methodology) = Modeling

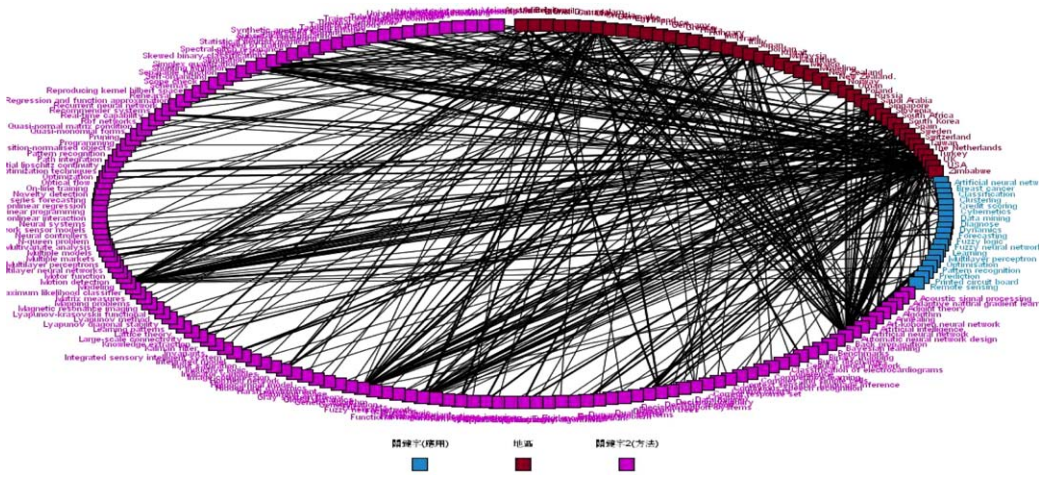

Fig. 4. Web diagram of keywords and first author's nationality association.

propagation' and 'modeling' are two methodologies highly associated with the first author's nationality. In addition, Asian countries such as 'China', 'Taiwan', 'Japan', 'Turkey', and 'India' prefer to use 'Back propagation' as the ANNs methodology. On the other hand, some non-Asian countries such as 'Switzerland', 'Brazil', 'Turkey', 'France', 'Spain', and 'UK' generally use 'Modeling' as methodology. More details are shown in Table 2.

3.3. Keywords and research category association rules

The research category categorizes retrieval articles as either science (S) or social science (SS) based on the Thomson ISI science and social science categories. In the databases, we found that only 1/5 of the articles were in the field of social science, with most of them in management and business studies, such as stock market analysis, foreign exchange forecasting, customer behavior research, and credit scoring. Thus, the vast majority of ANNs studies are in science research. Fig. 5 shows the association rules for keywords and research category.

There were four association rules extracted from mining results. In social science fields, only 'Genetic algorithm' is highly associated with consequent and antecedent. This implies that 'Genetic algorithm' has great likelihood to be implemented as a methodology for different social science applications. On the other hand, in the science research field, 'Back propagation', 'Modeling', and 'Fuzzy

Table 2
Association rules for keywords and first author's nationality (min sup. = 5, min conf. = 5)

	Support	Confidence	Lift	Consequent	Antecedent
1	8.50	5.80	11.77	Author's nationality = China	Keywords (application) = Global exponential stability
2	8.50	7.00	4.41	Author's nationality = China	Keywords (methodology) = Hopfield network
3	8.50	8.10	1.07	Author's nationality = China	Keywords (methodology) = Back propagation
4	7.00	5.60	4.39	Author's nationality = Japan	Keywords (application) = Learning
5	7.00	8.50	1.86	Author's nationality = Japan	Keywords (methodology) = Genetic algorithm
6	7.00	5.60	1.14	Author's nationality = Japan	Keywords (methodology) = Fuzzy logic
7	7.00	8.50	1.11	Author's nationality = Japan	Keywords (methodology) = Back propagation
8	6.20	6.30	16.06	Author's nationality = Taiwan	Keywords (methodology) = Credit scoring
9	6.20	6.30	16.06	Author's nationality = Taiwan	Keywords (methodology) = Taguchi method
10	6.20	11.10	2.25	Author's nationality = Taiwan	Keywords (methodology) = Fuzzy logic
11	6.20	9.50	1.25	Author's nationality = Taiwan	Keywords (methodology) = Hopfield models
12	10.10	6.90	1.18	Author's nationality = UK	Keywords (methodology) = Modeling
13	19.00	6.80	1.37	Author's nationality = USA	Keywords (methodology) = Fuzzy logic
14	7.60	7.80	2.32	Author's nationality = India	Keywords (methodology) = Back propagation
15	7.60	6.50	1.53	Author's nationality = Turkey	Keywords (methodology) = Back propagation
16	7.60	7.80	1.25	Author's nationality = Taiwan	Keywords (methodology) = Back propagation
17	7.60	7.80	1.11	Author's nationality = Japan	Keywords (methodology) = Back propagation
18	7.60	9.10	1.07	Author's nationality = China	Keywords (methodology) = Back propagation
19	5.80	5.10	12.86	Author's nationality = Switzerland	Keywords (methodology) = Modeling
20	5.80	5.10	3.03	Author's nationality = Brazil	Keywords (methodology) = Modeling
21	5.80	8.50	1.99	Author's nationality = Turkey	Keywords (methodology) = Modeling
22	5.80	5.10	1.77	Author's nationality = France	Keywords (methodology) = Modeling
23	5.80	5.10	1.39	Author's nationality = Spain	Keywords (methodology) = Modeling
24	5.80	11.90	1.18	Author's nationality = UK	Keywords (methodology) = Modeling

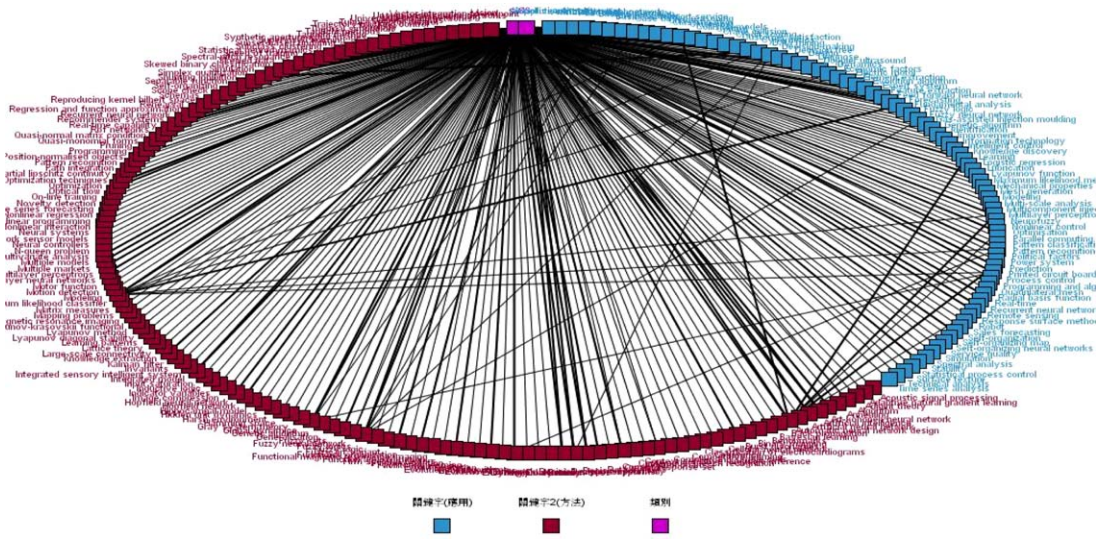

Fig. 5. Web diagram of keywords and research category association.

Table 3
Association rules for keywords and research category (min sup. = 5, min conf. = 5)

	Support	Confidence	Lift	Consequent	Antecedent	
1	7.20	9.60	2.26	Keywords (application) = Classification	Keywords (methodology) = Back propagation	Category = S
2	9.00	8.80	1.93	Keywords (methodology) = Genetic algorithm	Category = SS	
3	5.20	5.70	1.33	Keywords (application) = Classification	Keywords (methodology) = Modeling	Category = S
4	90.80	5.20	1.06	Keywords (methodology) = Fuzzy logic	Category = S	

logic' are the main methodologies for implementation. In addition, 'Back propagation' and 'Modeling' for methodologies integrated with the application of 'Classification' have a high lift value of 2.26. This means that these methodologies and application are highly implemented in ANNs science research (Table 3).

3.4. Keywords and year of publication association rules

The year of publication decision variable divided data into two groups as articles in the period from 1995 to 2000, and in the period from 2001 to present. The reason for choosing these periods was that the Internet was opened to general users in 1994 and most digital journals began online access in 1995. In this regard, this paper investigates whether there has been a difference in ANNs development in the most recent five years. Fig. 6 shows the association rules for keywords and year of publication.

Three association rules were found from the mining results. Articles published before year 2000 used the methodologies 'Fuzzy logic' and 'Back propagation'. On the other hand, only 'Modeling' is shown as an association rule

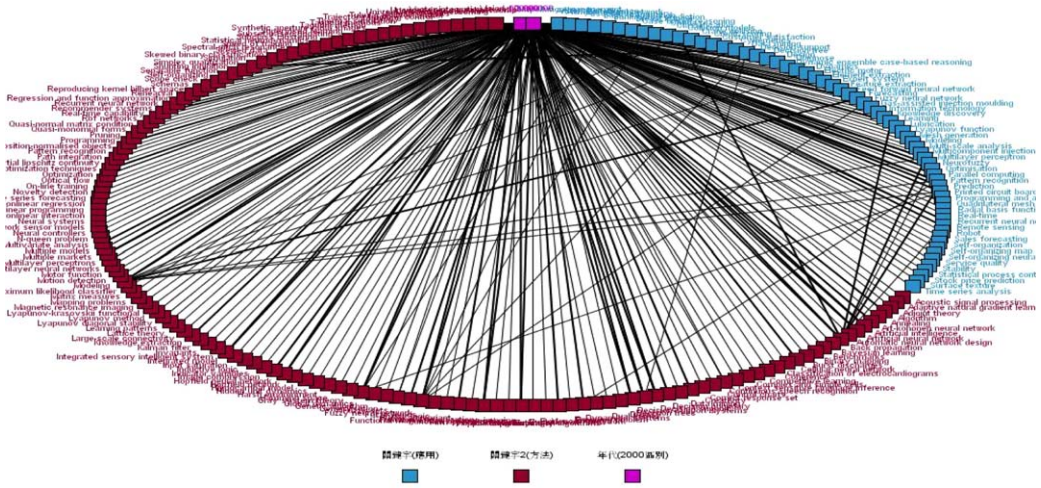

Fig. 6. Web diagram for keywords and year for publication association.

Table 4
Association rules for keywords and year of publication (min sup. = 5, min conf. = 5)

	Support	Confidence	Lift	Consequent	Antecedent
1	37.70	5.80	1.17	Keywords (methodology) = Fuzzy logic	Year = 1 (before 2000)
2	62.20	6.70	1.14	Keywords (methodology) = Modeling	Year = 2 (after 2000)
3	37.70	8.10	1.07	Keywords (methodology) = Back propagation	Year = 1 (before 2000)

for articles published after 2000 (Table 4). We tried to adjust the threshold value in order to extend applications on mining consequent, although no further association rules can be found.

4. Cluster analysis

Compared to association rule results, this research implements cluster analysis in this section in order to investigate whether there are different results between association rule and cluster analysis with the same decision variables.

4.1. Keywords and methodologies/applications cluster analysis

In Section 3.1, five association rules on keywords and methodologies/applications were presented. Thus, this research implements a self-automated clustering approach on five data groups to compare the analysis results with association rules. By doing so, mining results of cluster analysis on keywords and methodologies/applications are shown in Table 5.

Regarding methodologies, there are the five cluster sets of 'Genetic algorithm', 'Simulation', 'Artificial neural network', 'Back propagation', and 'Modeling'. Percentage number for each cluster set indicates the percentage of each keyword for different clusters and levels. In addition, the methodologies 'Back propagation' and 'Modeling', as well as the applications 'Classification' and 'Multilayer perception' show high percentages on different clusters and have similar results to association rules. On the other hand, the combination of the methodology 'Genetic algorithm' and the application 'Pattern recognition' has a high percentage for its cluster set. However, this finding does not exist in association rules and is different from the association rule analysis in Section 3.1. This means that different data mining methodologies present different results for keywords and methodologies/applications classification and clustering.

4.2. Keywords and author's nationality cluster analysis

Firstly, using the country as a decision variable for clustering, the mining results show four clusters, which contain the combination of country, methodology, and application including 'South Korea – Fuzzy logic – Data mining', 'China – Back propagation – Learning', 'USA – Back propagation – Multilayer perception', and 'Taiwan – Modeling – Artificial neural network'. In these clusters, only the cluster set of 'China – Back propagation – Learning' has the same result as the association rule in Section 3.2. USA and Taiwan have new clusters on ANNs methodologies and applications. On the other hand, South Korea is new cluster set that is not shown on association rule analysis (Table 6).

Secondly, when using keywords as decision variable for clustering, there is the same cluster set of 'Back propagation – Multilayer perception – China' as from the association rules in Section 3.2, and other cluster sets show new classification. For example, 'Genetic algorithm – Learning – Japan', 'Fuzzy logic – Artificial neural network – Taiwan', and 'Simulation – Cybernetics – USA'. Accordingly, different data mining methodologies also present different results for keywords and author's nationality classification and clustering (Table 7).

Table 5
Cluster sets of keywords and methodologies/applications

Instances	351	460	267	268	379
Keywords (methodology)	Genetic algorithm (49.57%)	Simulation (24.35%)	Artificial neural network (7.12%)	Back propagation (100%)	Modeling (52.51%)
Keywords (application)	Pattern recognition (2.56%)	Artificial neural network (2.39%)	Feed forward neural network (2.25%)	Classification, multilayer perception (2.61%)	Classification (1.85%)

Table 6
Cluster sets of keywords and author's nationality – country as decision variable

Instances	484	437	600	587
Country	South Korea (16.94%)	China (25.86%)	USA (45.33%)	Taiwan (16.18%)
Keywords (methodology)	Fuzzy logic (42.15%)	Back propagation (37.99%)	Back propagation (21%)	Modeling (33.22%)
Keywords (application)	Data mining (1.65%)	Learning (1.6%)	Multilayer perception (1.5%)	Artificial neural network (3.41%)

Table 7
Cluster sets of keywords and author's nationality – keywords as decision variable

Instances	447	634	559	468
Keywords (methodology)	Genetic algorithm (48.77%)	Back propagation (43.69%)	Fuzzy logic (23.79%)	Simulation (27.78%)
Keywords (application)	Learning (1.34%)	Multilayer perception (1.89%)	Artificial neural network (4.29%)	Cybernetics (2.56%)
Country	Japan (25.73%)	China (17.82%)	Taiwan (24.33%)	USA (58.33%)

4.3. Keywords and research category cluster analysis

Using the research category as the decision variable for clustering, in science research the same results are found as in association rule analysis for the classification of 'Modeling – Classification – Data mining'. However, 'Back propagation – Pattern recognition' and 'Genetic algorithm – Artificial neural network' are new classifications for the cluster set of keywords and research category, which is different from the results in the Section 3.3. On the other hand, in social science, the cluster set of methodologies and applications are 'Simulation', 'Artificial neural network', and 'Data mining', which are different from the association rule of 'Genetic algorithm'. Also, different data mining methodologies present different results for keywords and research category classification and clustering (Table 8).

4.4. Keywords and year for publication cluster analysis

Using the year of publication as the decision variable for clustering, three cluster sets were extracted by cluster analysis, showing a great difference from the association rules in Section 3.4. Compared with the association rules in Table 4, the similar mining result of association rule and cluster analysis is found in the data group from 1995 to 2000. On the other hand, 'Back propagation', 'Artificial neural network', and 'Learning' are a new cluster set that is different from the association rule for articles published from 2001 to 2005 in Table 4. This means that cluster analysis can find new classifications results for keywords and year of publication (Table 9).

5. Discussion

(1) Using keywords as literature review method has been evaluated in some articles (Liao, 2003, 2005). However, those classification methods had no deep analysis that compared different data groups with the same decision variables. This research implements a data mining approach including association rules and cluster analysis in order to analyze the ANNs literature through keyword classification and clustering of articles from 1995 to 2005 as a basis, exploring the ANNs methodologies and applications developments during that period. By doing so, this paper finds some association rules and cluster sets, that describe different patterns in relation to ANNs methodologies and applications. In addition, using different data mining approaches, this research examines for classification and clustering ANNs methodologies and applications with different mining results. This could be a literature analysis approach, which may overcome the limitations of data analysis using a single approach (Wong & Monaco, 1995; Wong & Selvi, 1998).

(2) This paper emphasis the importance using the keywords of academic articles or theses for research because these contain the three aspects of a research including research problem, theory, and methods/tools. In addition to the article title, by using the keywords of an article, its central theme can be grasped in order to understand if it relates to specific research needs. Therefore, most online databases provide search and advanced search functions using article

Table 8
Cluster sets of keywords and research category

Instances	537	626	512	433
Research category	Science (99.81%)	Science (100%)	Social science (59.57%)	Science (99.08%)
Keywords (methodology)	Back propagation (53.82%)	Genetic algorithm (26.68%)	Simulation (27.54%)	Modeling (50.35%)
Keywords (application)	Pattern recognition (3.35%)	Artificial neural network (2.72%)	Artificial neural network, data mining (2.34%)	Classification, data mining (1.85%)

Table 9
Cluster sets of keywords and year of publication

Instances	777	648	683
Year of publication	Before 2000 (79.54%)	Before 2000 (100%)	After 2001 (100%)
Keywords (methodology)	Fuzzy logic (27.03%)	Back propagation (28.09%)	Back propagation (17.57%)
Keywords (application)	Data mining (2.19%)	Classification (2.47%)	Artificial neural network, learning (1.32%)

keyword searches. This can help researchers reduce the search time on thousand of journals. Therefore, this research uses keywords as an index to investigate the development of ANNs by searching for articles on different online databases. We believe that digital format articles and online databases will be an increasingly important resource for research in the future. In this regard, by searching more online database a broader research scope and horizon can be obtained.

(3) This research presents data mining results, which describe association rules and cluster sets of ANNs development over a specific period. Some methodologies and applications are classified as research patterns, such as 'Back propagation – Classification', 'Modeling – Classification', 'Genetic algorithm – Pattern recognition', 'Back propagation – Multiplayer perception', and 'Simulation – Feed forward neural network'. This implies that most researchers implemented those methodologies and applications in their studies, and this could provide a basis for reviewing the development of ANNs on these research issues over the past decade. However, some advanced ANNs methodologies and their applications are not illustrated in this paper since it only provides a classification and clustering results for ANNs over a broad direction. On the other hand, integration of other methodologies with ANNs, such genetic algorithms and mathematical algorithms for modeling, are the trend of ANNs development. These algorithms can either reduce research limits on specific problem domain or increase capabilities for extending the integration of methodologies and their applications. Therefore, this paper suggests that integration of different methodologies and applications of ANNs could be an alternative research field.

(4) This research implements association rules and cluster analysis as a data mining approach. As we can see, mining results are partially consistent under the same decision variables. In this regard, this paper suggests that using more data mining approaches, such as decision tree, genetic algorithm, and even ANNs, could implement the exploration of whether different mining results could provide deeper explanations on the development of ANNs. We expect that different results of data mining on ANNs classification and clustering could broaden our horizons from other researches.

(5) Although 10,120 articles retrieved from five online databases were used for analysis, some other articles may have implemented similar ANNs methodologies in their applications without an ANNs index, so they may not have been found. Therefore, the first limitation of this article is the authors' limited knowledge in presenting an overall picture of this subject. Secondly, other academic journals/magazines, practical articles and reports are not included in this survey. These would have provided more complete information to explore the development of ANNs methodologies and applications. Thirdly, non-English publications were not considered in this survey to determine the effects of different cultures on the development of ANNs methodologies and applications. We believe that ANNs methodologies and applications in addition to those discussed in this article have also been developed and published in other areas and languages. Therefore, we prefer to consider that articles selected by this research are only part sampling from ANNs researches and this could be cause research errors on this research procedure and results.

(6) In addition, this research used SPSS Clementine data mining functions under the Windows XP environment with Intel Pentium 4 2.2 GHz CPU and 8 GHz RAM. However, the system has some limitation for data mining because a total of 110,080 data items were implemented on the data mining. Due to system limitations, this research could not integrate more decision variables and levels for data mining, and this may have limited the scope of classification and clustering results from the data mining process. Therefore, we suggest that more advanced IT infrastructures could provide enhanced mining results. In addition, different decision variables may provide different aspects for data mining. Finally, ANNs comprise a broad inter-disciplinary research topic. We expect that further research could improve research results and scope on ANNs development.

6. Conclusion

Based on a scope of 10,120 articles on ANNs, this paper uses data mining approach including association rules and cluster analysis to survey the ANNs through keyword classification and clustering based on articles from 1995 to 2005, exploring the ANNs methodologies and applications development during that period. The four decision variables of keywords, author's nationality, research category, and year of publication, with a total of 110,080 data items were implemented for data mining. This research shows that some specific ANNs methodologies and applications patterns could be extracted from mining results, which describe the ANNs development over this period. In addition, this paper finds that different classification and clustering results result from different data mining approaches. Therefore, using more data mining approaches for analysis could provide different explanations for ANNs development.

References

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of the 20th international conference on very large databases* (pp. 487–499).

Carpenter, G. A., & Grossberg, S. (1992). A self-organizing neural network for supervised learning, recognition, and prediction. *IEEE Communication Magazine, 30*, 38–49.

Coenen, F., Goulbourne, G., & Leng, P. (2004). Tree structures for mining association. *Rules, Data Mining and Knowledge Discovery, 8*, 25–51.

Hopfield, J. J. (1982). Neural systems and physical systems with emergent selective computational abilities. *Proceedings of Natural Academy Science in USA, 79*, 2554–2558.

Kouris, I. N., Makris, C. H., & Tsakalidis, A. K. (2005). Using information retrieval techniques for supporting data mining. *Data & Knowledge Engineering, 52*, 353–383.

Kuo, R. J., Liao, J. L., & Tu, C. (2005). Integration of ART2 neural network and genetic *K*-means algorithm for analyzing Web browsing paths in electronic commerce. *Decision Support Systems, 40*(2), 355–374.

Kohonen, T. (1988). *Self-organization and associative memory*. *Springer-series in information sciences* (Vol. 8). Berlin: Springer-Verlag.

Liao, S. H. (2002). Problem solving and knowledge inertia. *Expert Systems with Applications, 22*, 21–31.

Liao, S. H. (2003). Knowledge management technologies and applications – literature review from 1995 to 2002. *Expert Systems with Applications, 25*, 155–164.

Liao, S. H. (2005). Expert system methodologies and applications – A decade review from 1995 to 2004. *Expert Systems with Applications, 28*, 93–103.

Poggio, T., & Girosi, F. (1989). A theory of networks for approximation and learning. AI memo/CBIP paper 31, Massachusetts Institute of Technology, Cambridge, MA, p. 65.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536.

Schalkoff, R. J. (1997). *Artificial neural networks*. New York: McGraw-Hill.

Tsechansky, M., Pliskin, N., Rabinowitz, G., & Porath, A. (1999). Mining relational patterns from multiple relational tables. *Decision Support Systems, 27*, 177–195.

Vrahatis, M. N., Boutsinas, B., Alevizos, P., & Pavlides, G. (2002). The new *k*-windows algorithm for improving the *k*-means clustering algorithm. *Journal of Complexity, 18*(1), 375–391.

Wang, Y. F., Chuang, Y. L., Hsu, M. H., & Keh, H. C. (2004). A personalized recommender system for the cosmetic business. *Expert Systems with Applications, 26*, 427–434.

Wong, B. K., & Monaco, A. (1995). Expert system applications in business: A review and analysis of the literature (1977–1993). *Information & Management, 29*(3), 141–152.

Wong, B. K., & Selvi, Y. (1998). Neural network applications in finance: A review and analysis of literature (1990-1996). *Information & Management, 34*(3), 129–139.