Winter 12-10-2002

A Detailed Procedure for Using Copulas to Classify E-Business Data

Victor L. Berardi

B. Eddy Patuwo

Michael Y. Hu

A Detailed Procedure for Using Copulas to Classify E-Business Data

Victor L. Berardi, B. Eddy Patuwo, Michael Y. Hu
Graduate School of Management
Kent State University
Kent, Ohio 44224
epatuwo@bsa3.kent.edu

Abstract

Decision support systems are widely implemented to effectively utilize the tremendous amount of data generated by information systems throughout an organization. In one common implementation, the goal is to correctly classify a customer so that appropriate action can take place. This may take the form of a customized purchase incentive given to increase the probability that a transaction is completed, while enhancing profitability. Intelligent agents employing neural network technology that function as Bayesian classifiers are one approach used here.

Another approach that has been around for decades, called copulas, to our knowledge has yet to be utilized for classification in e-business applications. Copulas are functions that can describe the dependence among random variables. The very fact that copulas directly address codependence among variables may make them especially attractive in e-business applications where large numbers of correlated attributes may be present that could negatively affect the performance of other methods. In this paper, the basics of Bayesian decision making and posterior probabilities are reviewed. A detailed procedure for using copulas as Bayesian classifiers for e-business data is presented. The emphasis in describing the method is placed upon practitioner understanding to facilitate replication in real situations while maintaining technical rigor to ease computerized implementation.

1. Introduction

The impact of technology is being felt throughout the organization. Information technology is redefining not only how organizations develop strategies but the concepts of organizational learning and the nature of the organizational forms that are possible (e.g., [5] [14] [19]). At the tactical and operational levels, processes throughout the organization are rapidly being reengineered to incorporate the e-business approach.

All these information systems generate a tremendous amount of data for use in organizational decision processes. Effectively utilizing these data repositories is a key to not only meeting articulated business needs but to maximizing an organization's return on investment as well. For example, consider a common e-commerce application. Web-based retailers often track customer search and browsing records in addition to purchase information. This data can then be linked with internal databases, with external affiliate sources, and with readily available demographic data to provide sophisticated customer profiles. As a result, individualized purchase incentives, suggested add-on purchase items, customized advertising schemes and email distributions can quickly be created to increase sales, enhance profitability, and develop customer loyalty (e.g., [5} [11] [18] [20]).

The goal in situations like the aforementioned is to accurately classify the customer (object) in question so that appropriate action can take place. Many technologies have been utilized to effect classifications in e-business applications. Some, such as data mining, intelligent agents, and neural computing are relatively new. Others, such as logistic regression, have been widely used for decades. Many times these methods are used in concert. For example, an intelligent agent may be employed to customize a follow-up email for a new customer. As attribute data is collected on the new customer, a neural computing model is used to estimate the probability of the customer making subsequent purchases. The intelligent agent, in turn, enters an appropriate enticement into the email.

Often, the approach employed by the classification method tries to estimate the probabilities, called posterior probabilities, associated with an object belonging to each of the possible groups based upon observed attributes. The classifier, in its simplest form, assigns the object as belonging to the group to which it has the highest probability of belonging. This is known as Bayesian classification and is often employed with neural networks and logistic regression applications.

Another approach that has been around for decades, called copulas, to our knowledge has yet to be utilized for classification in e-business applications. Copulas are functions that can describe the dependence among random variables and have been used in applications such as risk analysis [7], the analysis of accident precursor events [23], and for aggregating expert opinions [10]. In each of these situations, correlated events and variables are likely. As will be seen, copulas can be used to develop the likelihood functions that are integral to estimating the posterior probabilities used in Bayesian classification and seem promising for e-business applications. The very fact that copulas directly address codependence among variables may make them especially attractive in this realm where large numbers of correlated attributes may be present that could negatively affect the performance of regression models and neural networks.

The remainder of this paper is structured as follows. The next section reviews the basics of posterior probabilities and Bayesian decision making. Then, three models that can function as Bayesian classifiers in e-business applications: logistic regression, neural networks, and copulas are discussed. The goal here is to illustrate how each model approaches the Bayesian classification problem in a distinctly different manner. Particular attention is paid to copulas as they are not as widely used as the other two in e-business applications. Next, a step-by-step process for building and using a copula-based approach in an e-business environment is detailed. The final section concludes the paper and discusses further considerations for interested readers.

2. Posterior Probability and Bayes

In a classification problem setting, the underlying population generating process characterizes the relationship between the input attributes and the output classes. This relationship defines what is called the *posterior probability* distribution. At this point, simply note that the posterior probability is a quantity giving the probability that an object belongs to a specified class after we have observed information related to the classification decision.

For example, consider a situation in which perhaps 50 percent of customers who begin the checkout process at a website actually complete the transaction. Without any information concerning a customer, the best we could say is that there is a 50 percent (prior) probability that the next customer will actually complete the transaction. However, after obtaining information on the customer, such as site visit history or previous purchases made, for example, we could modify our probability of transaction completion either up or down from 50 percent to reflect this newly obtained information. Appropriate action could then be taken to maximize expected sales and profits from the customer and transaction.

Posterior probability forms the basis of the well-known Bayesian classification theory [8]. An object's posterior probability is a quantity of great interest, since it allows us to make optimal decisions regarding the class membership of new data. Consider a case with J groups (w_j) that are described by N variables (x_n) where $P(w_j)$ is the prior probability that a randomly selected object belongs to group j. According to *Bayes rule*, if we obtain information contained in an observation $(x_1, \ldots, x_n)$, the prior probability will be revised into the posterior probability, $P(w_j \mid x_1, \ldots, x_n)$, that the given object $(x_1, \ldots, x_n)$ belongs to group j. That is,

$$P(w_j \mid x_1, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n \mid w_j) \cdot P(w_j)}{\sum_j f(x_1, x_2, \ldots, x_n \mid w_j) \cdot P(w_j)} \quad (1)$$

where $f(x_1, x_2, \ldots, x_n \mid w_j)$ is the conditional likelihood function for each group, w_j, and the denominator represents the unconditional joint distribution $f(x_1, x_2, \ldots, x_n)$. Bayes rule shows how observing the value of $(x_1, \ldots, x_n)$, the attributes for the object, changes the prior probability $P(w_j)$ to the posterior probability $P(w_j \mid x_1, \ldots, x_n)$ upon which the classification decision is based.

Furthermore, suppose that a particular $(x_1, \ldots, x_n)$ is observed and is to be assigned to a group. Let $\lambda_{ij}((x_1, \ldots, x_n))$ be the cost of misclassifying $(x_1, \ldots, x_n)$ to group i when it actually belongs to group j. Since $P(w_j \mid x_1, \ldots, x_n)$ is the probability that the object belongs to group j given attributes $(x_1, \ldots, x_n)$, the expected loss associated with assigning $(x_1, \ldots, x_n)$ to group i can be minimized by following the *Bayesian decision rule* for classification,

Decide w_k for $(x_1, \ldots, x_n)$ if $L_k(x_1, \ldots, x_n)$

$$= \min_{i=1,2,\ldots,M} \sum_{j=1}^{M} \lambda_{ij}(x_1, \ldots, x_n) P(w_j \mid x_1, \ldots, x_n) \quad (2)$$

A loss function of particular interest in the literature is known as the symmetrical or zero-one loss function. The zero-one loss function is specified as $\lambda_{ij}(x_1, \ldots, x_n) = 1$ for $i \neq j$, and 0 otherwise. This binary loss function assigns a zero cost to a correct decision while assigning a unit loss to all misclassifications. In this case, the Bayesian decision rule is to assign an object to the group associated with the maximum posterior probability:

Decide w_k for $(x_1, \ldots, x_n)$ if $P(w_k \mid x_1, \ldots, x_n)$

$$= \max_{j=1,2,\ldots,M} P(\omega_j \mid x_1, \ldots, x_n) \ . \quad (3)$$

This decision rule yields a minimum expected misclassification rate or, in other words, the maximum overall number of correct classifications in the long run.

3. Models for Estimating Posterior Probabilities

The above discussion clearly shows the important role of posterior probabilities in the Bayesian classification decision. It is readily apparent from Bayes rule, that accurately estimating the likelihood functions is a key to obtaining reliable posterior probability estimates for objects. Logistic regression, neural networks, and copula-based methods can all be used as Bayesian classifiers as discussed next.

3.1 Logistic Regression

Logistic regression attempts to estimate posterior probabilities and likelihood functions based on the Bernoulli distribution by utilizing a logistic response function of the general form

$$E(w) = \frac{\exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n)}{1 + \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n)}. \quad (4)$$

where $E(w \mid x_1, \ldots, x_n) = P(w \mid x_1, \ldots, x_n)$. The maximum likelihood estimates for the β 's are determined via iterative numerical search procedures. Logistic regression is readily available in statistical packages such as SAS or SPSS and quickly yields a model for given data with minimal computational delay. Logistic regression is well understood by practitioners and has enjoyed success in a wide range of classification problem settings. Neter, Kutner, Nachtsheim, & Wasserman [17] contains an excellent discussion of logistic regression for interested readers.

Logistic regression does have limitations worth noting. Perhaps the most significant one is the *a priori* specification of the logistic response function, which may restrict the utility of the regression to model in the presence of complex relationships. For example, correlated inputs are known to impact model performance. Also, performance degradation may be seen in situations where the input attributes variance-covariance matrix is not constant between groups. As a practical matter, often a variable selection process is used to overcome such problems but adds another dimension of complexity to this fixed-form approach.

3.2 Neural Networks

Neural networks take a fundamentally different approach to estimating posterior probabilities for classification. A neural network is a parallel system of interconnected computing elements called nodes. Information is processed via the interaction between nodes where knowledge is represented by the weights of the connections between them. Figure 1 shows a simple three-layer feedforward network representative of those used in this research. The first or lowest layer is called the input layer where external information enters the network while the last or top layer is called the output layer where the network produces the model solution. The middle layer or hidden layer provides the connections necessary for neural networks to identify complex patterns in the data. All nodes in adjacent layers are connected by acyclic arcs from the input layer to the hidden layer and from the hidden layer to the output layer [2].

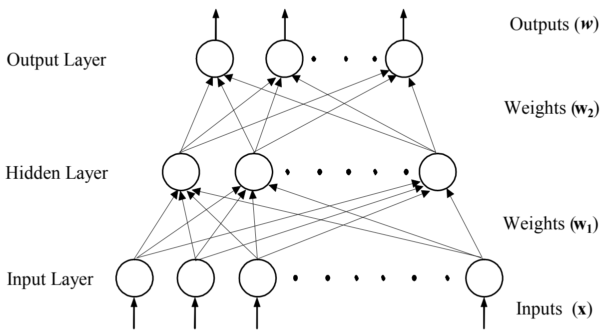

Figure 1. Multi-layer feedforward neural network.

An optimization process is utilized to minimize the difference between the targets—typically zero and one, representing the true groups to which the objects belong—and the neural network estimates of group membership. The theoretical relationship linking estimation of Bayesian posterior probabilities to neural networks has long been known (e.g., [1] [4] [9]). For an excellent discussion of designing and using neural networks for estimating posterior probabilities, interested readers are directed to [3].

Often neural networks are said to be parameter-free or data-dependent models. This means that the form of the neural network employed is not prespecified as in logistic regression. Rather, through experimentation with various numbers of hidden nodes, in conjunction with weight matrix determination via nonlinear optimization, the data itself determines the model form to be used. In this way, neural networks directly estimate the posterior probabilities associated with the observed objects and the likelihood functions they represent. This allows neural networks to model complex relationships in a wide array of problem settings and is a primary reason for their recent usage and success.

While neural networks are theoretically flexible they are not without practical limitation. The hidden node determination is not trivial, requiring multiple training and testing activities. In addition, the neural network optimization is sensitive to the initial weight values. This necessitates using multiple random initializations and optimizations at each hidden node to increase the probability of finding an acceptable solution. These factors combine to make model building quite time consuming and calculation intensive, even for less complex problems. Because the data are so integral to determining appropriate model form, with neural networks it is especially important to segregate some of the data to make selecting the best number of hidden nodes independent of neural network training.

3.3 Copulas

A copula approach to classification is fundamentally different from these other methods, and as such, holds potential as an alternative when other methods do not perform well or in combination with them. In general, the copula approach described in this paper attempts to decompose the joint cumulative distribution function represented by the data set into two components. First, the joint distribution is split into independent marginal cumulative distribution functions. Next, a second function, called a copula, is developed that specifies the interrelationship between the random variables. Then, these components are used to estimate the posterior probabilities associated with the objects in the data set. This copula approach to classification relies upon Skylar's Theorem, which is expressed as

Skylar's Theorem [22]. *Given a joint cumulative distribution function* $F(x_1,\ldots x_n)$ *for random variables* $X_1,\ldots,X_n$ *with marginal cumulative distribution functions (CDFs)* $F_1(x_1),\ldots,F_n(x_n)$, *F can be written as a function of its marginals:*

$$F(x_1,\ldots x_n) = C[F_1(x_1),\ldots,F_n(x_n)],$$

where $C(u_1,\ldots,u_n)$ *is a joint distribution with uniform marginals. Moreover, if each* F_i *is continuous, then C is unique, and if* F_i *is discrete, then C is unique on* $Ran(F_1)\times\cdots\times Ran(F_n)$ *, where* $Ran(F_i)$ *is the range of* F_i*.*

The function *C* is called a copula and describes the interrelationship between the random variables. Skylar's Theorem is especially useful because any joint distribution can be written in copula form without limitation.

To facilitate presentation and to ease understanding, consider the case where each F_i is continuous and differentiable, where the joint density $f(x_1,\ldots,x_n)$ is written as

$$f(x_1,\ldots,x_n) = f_1(x_1)\times\cdots\times f_n(x_n)c[F_1(x_1),\ldots,F_n(x_n)]\,. \quad (5)$$

$f_i(x_i)$ is the density that corresponds to $F_i(x_i)$ and $c = \partial^n C/(\partial F_1\cdots\partial F_n)$ and is known as the *copula density*. This shows that the joint density can be represented as a function of the marginal densities and the copula density [7]. For independent random variables, $c = 1$ indicates no dependence among the variables as the above equation simplifies to $f(x_1,\ldots,x_n) = f_1(x_1)\times\cdots\times f_n(x_n)$. When dependence among variables is present, $c \neq 1$ thereby encoding the interrelationship between variables. If a group w_j is specified, then $f(x_1,\ldots,x_n)$ becomes $f(x_1,\ldots,x_n \mid w_j)$ and the relationship to posterior probabilities via Bayes rule is clear.

Equation (5), therefore, provides a clear path on how to use the copula approach for estimating posterior probabilities in (1) for use in Bayesian classification per equations (2) and (3).

The copula density concept is likely to be unfamiliar to most readers, though, and deserves further discussion. The focus in this paper is on the multivariate normal copula family. It is called multivariate normal because it uses pairwise correlations between variables to represent dependence among variables in the same manner that the multivariate normal distribution is parameterized in terms of Pearson product-moment correlations [7]. By using a nonparametric correlation measure, such as the Spearman's ρ or Kendall's τ, however, multivariate normal copulas are flexible in that they can allow any marginal distribution for the input attributes. Steps 3 – 8 in the procedure below, detail precisely how this is accomplished.

As a result, the copula approach provides a flexible and useful approach to describing complex relationships among variables that can be used in conjunction with Bayes rule to develop Bayesian classifiers. Copulas have been underutilized given recent advances in the ease of data acquisition, the great speed of computational capabilities, and the availability of commercial packages enabling probability distribution fitting and estimation. Therefore, a primary goal of this paper is to detail a procedure for using a copula-based approach in decision making processes.

The copula approach to classification represents an exciting addition to the tool chest of techniques available for e-business applications. The novelty of its approach clearly distinguishes it from the other two. Copulas are by no means a panacea or guaranteed to perform better than the other methods. Copulas require estimates of marginal distributions and correlations between variables. This makes them more computationally intensive than logistic regression but less so than neural networks. In addition, many copula families exist and must be user specified. The multivariate normal copula family is particularly flexible and is used in the example problem as detailed in the procedure described next.

4. Procedure for Using Copulas with E-Business Data

In this section, a general procedure for developing and using multivariate copulas for classifying objects is presented. The goal is to clarify how copulas can be implemented with e-business data to improve decision making. The focus is on facilitating practioner understanding with an acknowledged tradeoff in brevity and computational efficiency. Readers interested in detailed theoretical development and proofs relating to copulas are directed to [7] [10] [16] [21.

There are four phases to using copulas in e-business applications. Phase one consists of collecting and processing the raw data. Phase two focuses on building the copula model. Phase three describes how to use the copula model to estimate posterior probabilities. Phase four, meanwhile, discusses using the copula model with newly acquired data.

4.1 Phase One—Collecting and Processing Data

STEP 1: Collect data on objects (e.g., customers) of interest from system.

STEP 2: Segment the objects into j groups (w_j) as appropriate for decision making.

4.2 Phase Two—Building a Copula Model to Couple the Marginal Distributions

STEP 3: Calculate correlation matrix ($\mathbf{R}_j$) for the independent variables $X_1,\ldots,X_n$ for each group w_j. While the multivariate normal copula is parameterized in terms of the Pearson product-moment correlations, to avoid the limitation of normally distributed independent variables, use nonparametric approaches such as Spearman's ρ or Kendall's τ thereby creating $\mathbf{R}_j^*$. Next, convert $\mathbf{R}_j^*$ into $\mathbf{R}_j$ via a simple transformation. For Spearman's, the conversion for each element r_{ij} is $r_{ij} = 2\sin(\pi\rho_{ij}/6)$. For Kendall's, the conversion is $r_{ij} = \sin(\pi\tau_{ij}/2)$. For example, in a two-group case with three input variables, two, 3x3 correlation matrices will be generated.

STEP 4: For each $\mathbf{R}_j$ calculate the inverse ($\mathbf{R}_j^{-1}$) and the square root of the determinant ($\sqrt{|\mathbf{R}_j|}$). These will be used for copula calculations in STEP 8.

STEP 5: Fit marginal distributions to each independent variable X_i for each group w_j. Save the distribution information for use in subsequent steps (e.g., Normal(μ, σ)). This can be accomplished using packages such as MatLab, SPSS, SAS, or using a distribution fitting program such as ExpertFit [12]. Law and Kelton [13], Morgan and Henrion [15], and Clemen [6] provide more information on this issue for interested readers. For a two-group three-input case, each of the three variables will be fitted twice.

STEP 6: For each object (realization) $(x_1^k,\ldots,x_n^k)$, compute the cumulative density functions (CDFs) and probability density functions (PDFs) associated with each of the marginal distributions found in STEP 5. Call the CDFs $(F_{11}^k, F_{12}^k,\cdots,F_{1n}^k)$ … $(F_{j1}^k, F_{j2}^k,\cdots,F_{jn}^k)$ for each of the j groups and n variables. Call the PDFs $(f_{11}^k, f_{12}^k,\cdots,f_{1n}^k)$ … $(f_{j1}^k, f_{j2}^k,\cdots,f_{jn}^k)$. Therefore, for a two-group case, each object will have two CDFs and two PDFs estimated for each variable.

STEP 7: For each of the k objects, compute the inverse univariate standard normal from the n CDFs found in STEP 6 as $y_{jn}^k = \Phi^{-1}(F_{jn}^k)$ where Φ ~ Normal(0,1). Let $\mathbf{y}_j^k$ be the column vector $\mathbf{y}_j^k = [y_{j1}^k, y_{j2}^k,\ldots,y_{jn}^k]$ of object k for group w_j. For a two-group case, each object will have two estimates, $\mathbf{y}_1^k$ and $\mathbf{y}_2^k$.

STEP 8: Calculate the point estimate of the copula density of each object k for every group w_j. The copula density is found as

$$C_j^k = e^{-\frac{1}{2}\{\mathbf{y}_j^{k\,t}(\mathbf{R}_j^{-1}-\mathbf{I})\cdot\mathbf{y}_j^k\}} / \sqrt{|\mathbf{R}_j|}$$

where $\mathbf{y}_j^{k\,t}$ is the transpose of $\mathbf{y}_j^k$.

4.3 Phase Three—Using the Copula Model to Estimate Posterior Probabilities

STEP 9: *For each object (realization) $(x_1^k,\ldots,x_n^k)$, compute the likelihood function for every group w_j.* The likelihood function is calculated as

$$f^k(x_1,x_2,\ldots,x_n \mid w_j) = f_{j1}^k(x_1)\times f_{j2}^k(x_2)\times\cdots\times f_{jn}^k(x_n)\times C_j^k.$$

STEP 10: Compute the unconditional joint distribution for each object. This is found as

$$f^k(x_1,x_2,\ldots,x_n) = \sum_{j=1}^{J} f^k(x_1,x_2,\ldots,x_n \mid w_j)\cdot P(w_j)$$

where $P(w_j)$ is the prior proportion of group w_j objects in the data.

STEP 11: Compute the posterior probabilities of each object belonging to each group. The posterior probabilities are calculated as

$$P^k(w_j \mid x_1,x_2,\ldots,x_n) = \frac{f^k(x_1,x_2,\ldots,x_n \mid w_j)\cdot P(w_j)}{f^k(x_1,x_2,\ldots,x_n)}$$

These probability estimates are then used in Bayesian decision making for classification as shown in equations (2) and (3). When an equal cost of misclassification assumption is used, the objects are assigned to the group corresponding to the highest posterior probability. Make appropriate decisions for each classification, such as offer a purchase enticement to objects classified as group w_2, a customer unlikely to purchase again.

The copula model has just been finalized and utilized to calculate posterior probabilities associate with the known objects. Next, we turn to the issue of how to use a copula model for classifying future observations.

4.4 Phase Four—Using Copula Model with New Data

STEP 12: Collect additional data on objects (e.g., customers) of interest from the system to be used to make immediate classification decisions and to take appropriate actions.

STEP 13: Repeat STEPS 6 – 11 for each newly obtained object. Use information calculated previously from STEPS 3 –5, but do not recalculate these steps. Make appropriate decisions in STEP 11 based upon the group to which the object is assigned.

5. Conclusion

Information systems are generating a tremendous amount of data for use in e-business decision processes. Effectively utilizing these data repositories to make decisions consistent with the organizations tactical and strategic objectives can be a key to achieving and sustaining competitive advantage. E-business decision support systems have frequently employed data mining, neural computing, intelligent agents, and regression approaches, alone or in concert to turn these data into information and knowledge.

Copulas are another approach that is promising for e-business applications. Like neural networks or logistic regression, copulas can be used to accomplish Bayesian decision making as part of a decision support system. Fundamentally, though, the copula approach to analyzing the data is quite different from these other methods. In particular, a copula approach directly addresses the codependence among variables. This may be especially desirable in e-business applications where large numbers of correlated variables exist that may negatively impact the performance of other methods.

In this paper, the basics of Bayesian decision making and posterior probabilities are reviewed. Next, the neural networks, logistic regression, and copula approaches to estimating the posterior probabilities upon which Bayesian decision making is based in discussed. The focus is on understanding how each method works so that the differences can be understood. Finally, a detailed method for using copulas to classify e-business data is presented. The emphasis is placed upon practitioner understanding to facilitate replication in real situations while maintaining technical rigor to ease computerized implementation.

6. References

[1] Baum, E.B., & Wilczek, F. Supervised learning of probability distributions by neural networks. In D. Anderson (Ed.) *Neural Information Processing Systems*, New York: American Institute of Physics, 1988, 52-61.

[2] Berardi, V.L., & Zhang, G.P. The effect of misclassification costs on neural network classifiers. *Decision Sciences*, 1999, 30, 13, 659 – 682.

[3] Bishop, C.M. *Neural Networks for Pattern Recognition*, Oxford, Clarendon Press, 1995.

[4] Bourlard, H., & Wellekens, C.J. Links between Markov models and multilayer perceptrons. In D.S. Toretzky (Ed.) *Advances in Neural Information Processing Systems*, San Mateo, CA: Morgan Kaufmann, 1989, 1, 502-510.

[5] Bult, J.R., & Wansbeek, T. Optimal selection for direct mail. *Marketing Science*, 1995, 14, 4, 378-394.

[6] Clemen , R.T. *Making Hard Decisions: An Introduction to Decision Analysis*, 2nd ed., Duxbury, Belmont, CA, 1996.

[7] Clemen, R.T. &Reilly, T. Correlations and copulas for decision and risk analysis. *Management Science*, 1999, 45, 2, 208-224.

[8] Duda, R.O., & Hart P. *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[9] Hung, M.S., Hu, M.Y., Patuwo, B.E., & Shanker, M. Estimating posterior probabilities in classification problems with neural networks. *International Journal of Computational Intelligence and Organizations*, 1996, 1, 49-60.

[10] Jouini, M., & Clemen, R.T. Copula models for aggregating expert opinions. *Operations Research*, 1996, 44, 3, 444-457.

[11] Karauga, G.G., Khraban, A.M., Nair, S.K., & Rice, D.O. AdPalette: an algorithm for customizing online advertisements on the fly. *Decision Support Systems*, 2001, 32, 85-106.

[12] Law, A.M. ExpertFit, Averill M. Law & Associates, www.averill-law.com, 2002.

[13] Law, A.M. and W.D. Kelton. *Simulation Modeling and Analysis, 2nd Edition*. McGraw-Hill, 1991.

[14] Li, Shuliang. The Development of a Hybrid Intelligent System for Developing Market Strategy, *Decision Support Systems,* 2000, vol.27, pp.395-409.

[15] Morgan, G.M., & Henrion, M. *Uncertainty*. Cambridge University Press, Cambridge, 1990.

[16] Nelson, R.B. Copulas, characterization, correlation, and counterexamples, Mathematics Magazine, 1995, 68, 193-198.

[17] Neter, J., Kutner, M.H., Nachtsheim, C.J., & Wasserman, W. *Applied Linear Statistical Models, 4th Edition*, McGraw-Hill, 1996.

[18] Palmer, Ann T. Online Savings Are In Bloom, *Business Week*, July 22, 2002. pg. 10.

[19] Raymond, L., & Blili, S. Organizational learning as a foundation of electronic commerce in the network organization. *International Journal of Electronic Commerce*, 2001, 5, 2, 29-45.

[20] Schmittlein, D.C., & Peterson, R.A.. Customer base analysis: an industrial purchase process application. *Marketing Science,* 1994, 13, 1, 41-67.

[21] Schweizer, B. Thirty years of copulas. G. Dall'Aglio, S. Kotz, G. Salinetti, eds. *Advances in Probability Distributions with Given Marginals*, Kluwer, Dordrecht, Netherlands. 1991, 13-50.

[22] Skylar, A. Fonctions de repartition a *n* dimensions et leurs marges. *Publications de l'Institut Statistique de l'Universite de Paris*, 1959, 8, 229-231.

[23] Yi, W., & V. m. Bier. An application of copulas to accident precursor analysis. *Management Science*, 44, S257-S270, 1998.