

A Protocol and Tool for Developing a Descriptive Behavioral Model

by

Steven Jeck

April 2012

Chair: Karl Abrahamson, PhD.

Major Department: Computer Science

Fuzzy rules have been used to model complex human behavior in order to develop sophisticated industrial control systems. The use of fuzzy rules to create a behavioral model provides a quantitative basis for discussing the contribution of elements of the model to theories about the behavior. The application of a protocol and tool simplifies the development of a behavioral model from observational data. Extraction of a high level, linguistic behavioral model from the observational data is used to discover knowledge about the data. Tuning of the model is accomplished by parameter optimization through the adjustment of membership functions using the genetic fuzzy, self-adaptive system. A case study demonstrating the use of the protocol and tool is presented. In the study, a behavioral model is developed that integrates the analysis of the observational data with Social Network Analysis. The integrated behavioral model provides an effective platform for a quantitative analysis of the activities impacting behavior.

A Protocol and Tool for Developing a Descriptive Behavioral Model

A Thesis

Presented to

the Faculty of the Department of Computer Science

East Carolina University

In Partial Fulfillment

of

the Requirements for the Degree

Master of Science in Software Engineering

by

Steven Jeck

April 2012

Copyright 2012, Steven Jeck

A Protocol and Tool for Developing a Descriptive Behavioral Model

by

Steven Jeck

APPROVED BY:

DIRECTOR OF THESIS: _____ John Placer, PhD

COMMITTEE MEMBER: _____ Nasseh Tabrizi, PhD

COMMITTEE MEMBER: _____ Qin Ding, PhD

COMMITTEE MEMBER: _____ Jennifer Verdolin, PhD

CHAIR OF THE DEPARTMENT OF
COMPUTER SCIENCE: _____ Karl Abrahamson, PhD

DEAN OF THE GRADUATE
SCHOOL: _____ Paul J. Gemperline, PhD

TABLE OF CONTENTS

1	INTRODUCTION.....	1
1.1	Outline of the Paper	5
1.2	Developing a Behavioral Model from Data: A Case Study.....	5
1.3	Investigating Relationships with the Behavioral Model: A Case Study	5
1.4	Conclusions	5
2	DEVELOPING A BEHAVIORAL MODEL FROM DATA: A CASE STUDY.....	6
2.1	Defining the parameters of the Behavioral Model	8
2.2	Defining the structure of the Behavioral Model	11
2.3	Extracting a Descriptive Behavioral Model	13
2.4	Classifying Individuals with the Behavioral Model.....	24
3	INVESTIGATING RELATIONSHIPS WITH THE BEHAVIORAL MODEL: A CASE STUDY.....	25
3.1	Chromosome Encoding.....	26
3.2	Fitness Function.....	30
3.3	Reproduction, Mutation and Natural Selection	34
3.4	Application to the Case Study	37
3.4.1	Complete Data Set Tests.....	40
3.4.2	Group Level Tests	43
4	CONCLUSIONS.....	45

LIST OF TABLES

Table 2-1 – The Membership Function File	10
Table 2-2 – Classification Type Definition	11
Table 2-3 – Preliminary Rule	13
Table 2-4 – Statistical Analysis used by Analyze Data	15
Table 2-5 – Analyze Values output for TIMID Rules	18
Table 2-6 – Analyze Values output for NORMAL Rules	20
Table 2-7 – Analyze Values output for BOLD Rules	22
Table 2-8 – High level statements about the classification types	23
Table 2-9 – Evaluate Rules output	24
Table 3-1 – Constraint test for valid trapezoid	29
Table 3-2 – Comparison of Fitness Evaluation Methods	38
Table 3-3 – Comparison of population size test results	39
Table 3-4 – Correlation Test Results	41
Table 3-5 – Betweenness Test Results	43
Table 3-6 – Degrees Test Results	44
Table 3-7 – Path Length Test Results	44
Table 3-8 – Total Interactions Test Results	44

LIST OF FIGURES

Figure 2-1 - BMA Architecture	7
Figure 2-2 – Membership Functions for a Fuzzy Variable	9
Figure 2-3 – Analyze Values output for All Rules	15
Figure 2-4 – Analyze Values output for TIMID Rules.....	17
Figure 2-5 – Analyze Values output for NORMAL Rules	19
Figure 2-6 – Analyze Values output for BOLD Rules.....	21
Figure 3-1 – Initial random membership function definition	28
Figure 3-2 – Initial randomly defined gene	30
Figure 3-3 - Gene level crossover	35
Figure 3-4 – Trappability Membership Function Comparison	42

1 Introduction

Researchers in the social and biological sciences collect data based on the observation of individual and group activities. With the decreased cost of video and computer monitoring equipment, greater volumes of data are available than ever before. However, even with larger sets of observational data, and perhaps, due to the volume of data, understanding the data is increasingly difficult. Data mining techniques have been employed to help businesses discover relationships in the data often leading to unusual discoveries about the types of products purchased [Wasserman]. Classic data mining relies on statistical analysis of the database to discover association. While providing useful information, these techniques do not support the researcher's investigation into theories of behavior and the factors that lead to the decisions.

Methods have been developed for classifying animal behavior but are generally limited in the use of social context. [Wilson] measures "personality" based on the willingness of an individual to take risks when presented with a stimulus. This type of model does not account for the impact of the activities of others or the time constraints an individual's activities impose on their own behavior. Social Network Analysis has been used by biologists to study the communication between individuals in a group [Krause] [Lusseau]. The use of Social Network Analysis in understanding social dynamics improves the ability to understand the interactions of individuals within a group [Sih]. No work has been found in the literature which provides a method of developing a behavioral model that integrates the analysis of observational data with Social Network Analysis. An integrated behavioral model would provide a quantitative

basis for discussing the activities that impact behavior. This thesis provides a technique to develop a behavioral model that can be used to evaluate the effect of an individual's activities on their Social Network Analysis metrics.

Fuzzy rules have been used to model complex human behavior in order to develop sophisticated industrial control systems. Fuzzy Rule Based Systems (FRBS) are typically used in control systems where complex input cannot be easily reduced to an algorithmic output value. FRBS rely on the encoding of heuristic information to process input data and generate output. These systems use linguistic variables and fuzzy rules to model the behavior of an expert operator to achieve an acceptable result. The use of linguistic variables results in an interpretable behavioral model expressed at a conceptual level easily understood by researchers without a deep understanding of fuzzy set theory. However, the definition of the membership functions used to map the linguistic variables and the fuzzy rules requires expert knowledge of the problem domain and is one of the more difficult tasks facing a FRBS designer [Berkan]. Despite the benefits of linguistic modeling, construction of a FRBS can be difficult. Extracting the expert knowledge necessary for constructing the rules is difficult if the process being modeled is not mature.

Evolutionary computing algorithms model the processes found in biological systems. [Holland 1975, 1992] described the concept of using biological systems and the Genetic Algorithm (GA) as the basis for computational algorithms. The GA is based on the processes found in biological systems and they are used in optimization of high dimensional problems. The coupling of a FRBS and GA to produce Genetic Fuzzy Systems (GFS) has proven successful in system identification methodologies [Cordón].

The GA has also found application in the discovery of fuzzy rules [Kim]. Genetic tuning is a commonly used technique where a GA is used to adjust the parameters of a defined FRBS. Common targets for genetic tuning are scaling functions, the universes of discourse and the membership functions. Genetic tuning is a parameter estimation approach in fuzzy system identification using GFS. Early efforts at using GAs to tune the FRBS were focused on optimizing the membership functions. [Karr] was one of the first to propose the use of a GA to modify the shapes of the membership functions.

This thesis proposes an approach not found in the literature: the application of FRBS techniques used in the development of industrial control systems to the development of a behavioral model for biological research. By creating a behavioral model consisting of a structure of fuzzy rules and parameters defined by the membership functions, a quantitative basis for discussing behavior and its classification is established. Using the quantitative behavioral model, the relationship between the activities of the individual in its social setting and the measure of those activities using Social Network Analysis can be investigated.

This thesis develops a protocol and a set of tools that can be used to create a descriptive behavioral model extracted from data obtained through the observation of individuals in a group. The protocol and tool define a procedure for investigating the relationship between the classification produced by the behavioral model and another metric of activity. A software program, the Behavior Model Analyzer (BMA), was designed and constructed to support the development and execution of the behavioral model as a classification system.

The protocol is a procedure used in conjunction with BMA to transform data obtained from the observation of activities into a set of fuzzy rules and fuzzy inputs that provide the structure and parameters of the behavioral model. BMA provides a tool to extract a high level, descriptive model from the fuzzy rules. Using BMA and the behavioral model, the data obtained from observing an individual's activities is used to classify the individual.

The BMA supports the investigation of the relationship between the classifications generated by the behavioral model and another measure of behavior. In the case study, the relationship between Social Network Analysis metrics of the individuals in a group and the behavioral model classification of the individuals were studied.

The following chapter provides an outline for the remaining sections of the thesis.

1.1 Outline of the Paper

This thesis is divided into four sections. Chapter 2 describes protocol for developing a behavioral model from observational data. Chapter 3 describes use of the BMA to investigate relationships between the classification index generated by the behavioral model and another measure of behavior. . Chapter 4 presents the conclusions from the use of the BMA and suggestions about future work.

1.2 Developing a Behavioral Model from Data: A Case Study

The second chapter describes the protocol for extracting fuzzy rules from observational data. A case study conducted in cooperation with a biology researcher studying animal behavior in a group is used to illustrate the protocol. The chapter includes a detailed description of the development of a behavioral model based on the protocol.

1.3 Investigating Relationships with the Behavioral Model: A Case Study

The third chapter describes the protocol for investigating relationships between the classifications provided by the behavioral model and the Social Network Analysis of the group of animals in the biology researcher's study.

1.4 Conclusions

The fifth chapter presents conclusions about the use of BMA and thoughts about future work in the development of behavioral models based on the protocol.

2 Developing a Behavioral Model from Data: A Case Study

The case study involves the use of the Behavior Model Analyzer (BMA) to assist a biology researcher interested in understanding how the activities of the individuals in a group of animals may be related to other measurements of activity. The researcher wanted to create a behavioral model that could be used to classify the individuals in a group and provide a numerical index for the classification. Having constructed the behavioral model, the researcher was interested in determining if there was a relationship between the classification provided by the behavioral model and another measurement of behavior such as Social Network Analysis.

The observational data used in the case study was obtained from a behavioral study conducted on Gunnison's prairie dogs in Flagstaff, Arizona. Data was collected from 3 plots on two separate colonies where individual prairie dogs were marked and their behavior was captured using video recording [Verdolin]. The behavioral data included behaviors such as interactions between individuals and time spent engaged in the behaviors of interest.

BMA is a Genetic Fuzzy System (GFS) constructed to support a protocol for the following:

- creation and development of a descriptive behavioral model
- the use of the behavioral model in the classification of individuals
- and the investigation of the relationships between the behavioral model classification and another measure of behavior.

The structure of the behavioral model is represented by the fuzzy rules of the Fuzzy Rule Based System (FRBS) component of BMA. The parameters of the behavioral model are the fuzzy inputs to the FRBS. Using the Genetic Algorithm (GA) component of BMA to adjust the parameters of the behavioral model, BMA seeks to find a relationship between the behavioral model classification of the individual and another measure of behavior.

The figure below shows the architecture of BMA.

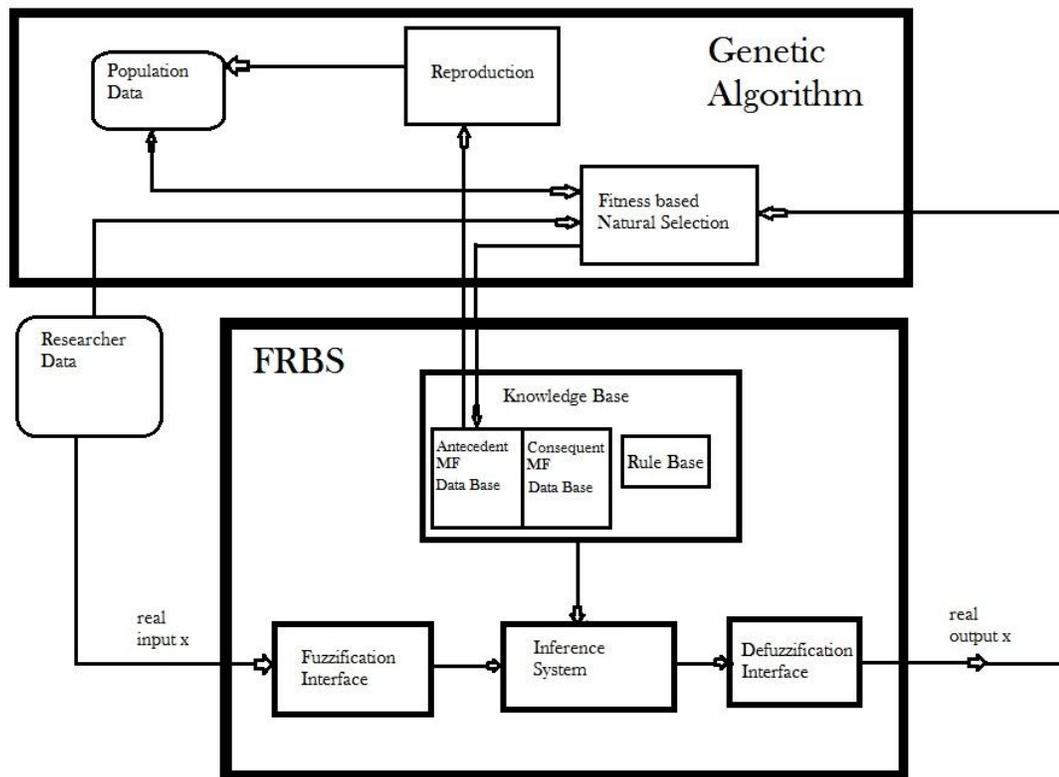


Figure 2-1 - BMA Architecture

A protocol was defined for the use of BMA. The first step in the protocol is to define the parameters of the behavioral model. The second step is to define the

structure of the behavioral model. The third step is to extract a high level, linguistic description of the behavioral model. The final step is to classify the individuals using the behavioral model.

2.1 Defining the parameters of the Behavioral Model

Defining the parameters is the first step in creating the behavioral model. The researcher begins the process by gathering data about the individuals for which the behavioral model will be developed. A fuzzy variable is defined to represent each of the activities being measured. The fuzzy variables are the parameters of the behavioral model. For each activity represented by a fuzzy variable, a set of fuzzy values must be defined. A fuzzy value is defined by a linguistic label and a membership function. The fuzzy value represents a conceptual category, such as SLOW, MEDIUM, and FAST, over the actual values a measurement may take. The measurement of an observed activity is a crisp value that is transformed by the membership function into a fuzzy value. Fuzzy rule based systems for mature processes such as industrial control systems may use many fuzzy values for each fuzzy variable [Cordón]. Since the understanding of animal behavior is more rudimentary, BMA limits each fuzzy variable to one of three possible fuzzy values. The figure below shows a typical set of membership functions used to represent the fuzzy values of a fuzzy variable.

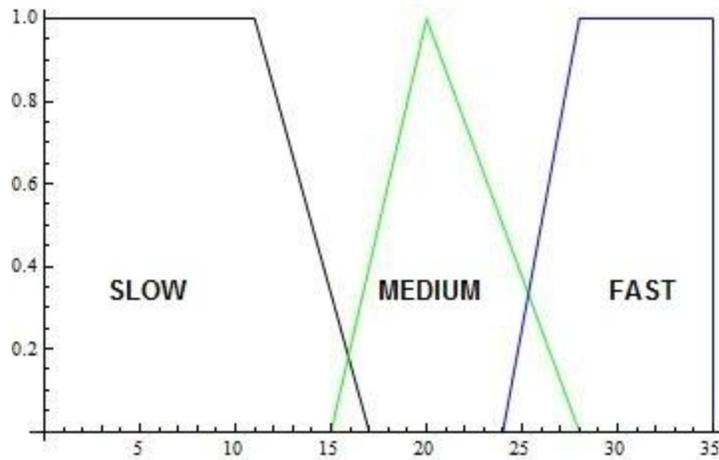


Figure 2-2 – Membership Functions for a Fuzzy Variable

The measurements may consist of crisp values that are not well represented by fuzzy values. For example, the sex of an individual is binary value, either male or female, yet the sex of the individual may have a significant impact on the rules in the behavioral model. During initialization, BMA handles instances of binary variables by matching the linguistic label used in the *Data* file, such as MALE, to a fuzzy value defined in the *Membership Function* file and selecting a crisp value for MALE that results in a membership value of 1.0.

In the case study, a *Membership Function* file containing 9 fuzzy variables and their associated fuzzy values was used. The *Membership Function* file is shown in the table below.

```
;Fuzzy membership functions
```

```
sex
```

```
=M, 0, 0, 0.4, 0.5
```

```
=U, 0.4, 0.49, 0.51, 0.6
```

```
=F, 0.51, 0.6, 1.0, 1.0
```

```
age
```

```
=YR, 0, 0, 0.8, 1
```

```
=U, 0.8, 1.0, 1.1, 1.5
```

```
=AD, 1.1, 1.5, 8, 8
```

```
trappability
```

```
=LOW, 0, 0, 20, 25
```

```
=MED, 20, 25, 55, 60
```

```
=HIGH, 55, 60, 100, 100
```

```
feeding
```

```
=LOW, 0, 0, 20, 25
```

```
=MED, 20, 25, 50, 55
```

```
=HIGH, 50, 55, 100, 100
```

```
vigilance
```

```
=LOW, 0, 0, 25, 30
```

```
=MED, 25, 30, 55, 60
```

```
=HIGH, 55, 60, 100, 100.0
```

```
aggression
```

```
=LOW, 0, 0, 2, 2.5
```

```
=MED, 2, 2.5, 5, 5.5
```

```
=HIGH, 5, 5.5, 100, 100
```

```
gk
```

```
=LOW, 0, 0, 2.5, 3
```

```
=MED, 2.5, 3, 5, 6
```

```
=HIGH, 5, 6, 100, 100
```

```
; Personality is the antecedent
```

```
personality
```

```
=TIMID,0, 0, 20, 30
```

```
=NORMAL, 20, 30, 80, 90
```

```
=BOLD, 80, 90, 100, 100
```

Table 2-1 – The Membership Function File

The modal points of the membership functions are initially defined by the researcher based on expert knowledge of the problem domain.

2.2 Defining the structure of the Behavioral Model

Defining the structure is the second step in constructing the behavioral model. A set of fuzzy rules define the structure of the behavioral model. Fuzzy rules are used to classify behavior by defining an output for a specific set of input values. BMA limits the fuzzy output to one fuzzy variable which may take one of three fuzzy values.

The classification type is based on the fuzzy output for the individual. The table below defines the classification type for a typical fuzzy variable with three fuzzy values of LOW, MEDIUM and HIGH.

Output Fuzzy Value	Classification Type
LOW	1
MEDIUM	2
LOW and MEDIUM	3
HIGH	4
LOW and HIGH	5
MEDIUM and HIGH	6
LOW and MEDIUM and HIGH	7

Table 2-2 – Classification Type Definition

Individuals with classification types 3 or 6 are examples of an individual with partial membership in both the LOW and MEDIUM or MEDIUM and HIGH fuzzy values. The occurrence of classification types 5 and 7, while possible, may also indicate issues with the rule set, since their occurrence together is at odds with a clear classification of the individual.

Creation of the fuzzy rules for the behavioral model requires expert understanding of the behavior being modeled. Expert understanding comes from

experience with the subject being modeled and analysis of observational data. BMA provides tools to enable an expert in the problem domain to rapidly develop a rule set by analyzing the data and producing a preliminary rule set.

The BMA *Create Rules* tool produces a preliminary rule set based on fuzzy values that occur in the data. For each individual in the *Data* file, the *Create Rules* tool converts the crisp values into fuzzy values using the membership functions defined in the *Membership Function* file. The fuzzy values are substituted into the Rule Template to generate a preliminary fuzzy rule. The consequent of the preliminary fuzzy rule is left blank. Defining the consequent's fuzzy value requires expert knowledge not contained in a general purpose tool like BMA. After all of the individuals are evaluated, a CREATE_RULES text file is generated containing all of the unique preliminary fuzzy rules. As some individuals may have common characteristics, fuzzy rule set consistency and redundancy is verified before the CREATE_RULES text file is generated. An expert in the problem domain must use their knowledge to define the fuzzy output value for each rule. The CREATE_RULES file will only be as complete as the data provided as input. The expert may choose to add, modify or delete rules from the preliminary rule set until a final rule set is defined.

Using the case study *Data* file, the *Create Rules* tool was used to generate a preliminary rule set. The rule set included 120 unique rules. One of the preliminary rules is shown in the table below.

IF sex IS F AND age IS YR AND trappability IS LOW AND feeding IS MED AND vigilance IS LOW AND moving IS HIGH AND resting IS LOW AND aggression IS MED AND gk IS LOW THEN PERSONALITY IS
--

Table 2-3 – Preliminary Rule

After the fuzzy output of each preliminary rule has been assigned by an expert in the problem domain, a behavioral model has been completely defined. The antecedents of the fuzzy rule are connected with the conjunction operator, AND. The use of a linguistic hedge in the fuzzy rule is not supported by BMA. Larson product implication is used to calculate the fuzzy output value for the fuzzy rule. Center Of Area (COA) is used to combine the output of the fuzzy rules into a crisp number representing a classification index for the behavioral model.

2.3 Extracting a Descriptive Behavioral Model

While each fuzzy rule clearly defines a behavior classification in linguistic terms, the fuzzy rule set can be quite large making the behavioral model difficult to interpret. BMA provides the *Analyze Values* tool to extract a high level, descriptive behavioral model from the fuzzy rules.

The *Analyze Values* tool provides a high level view of the fuzzy values used in the Rules file. Four different views of the rules are analyzed. The first analysis covers

all of the rules. Each of the remaining analyses reflects the use of antecedent fuzzy values for each of the different fuzzy values used in the Consequent. The *Analyze Values* tool generates two reports for each view. The first report is a bar chart representing the distribution of antecedent fuzzy values as they are used with the fuzzy variables in the rule set. The second report is a linguistic description of the distribution of fuzzy values as they are used with the fuzzy variables in the rule set. .

The linguistic description provides a high level description of the behavioral model in non-mathematical terms. The *Analyze Values* tool performs a statistical analysis of the data presented in the bar chart to determine how to describe the use of a fuzzy value. The mean and standard deviation are calculated for the use of fuzzy values for each fuzzy variable. The percentage of use each fuzzy value receives is compared against the mean and standard deviation to select the appropriate descriptive text. The table below describes the text selection criteria.

Criteria	Descriptive Text
>Mean + StdDev	STRONG emphasis on the use in rules
>Mean	SLIGHT emphasis on the use in rules
=Mean	AVERAGE emphasis on the use in rules
>Mean – StdDev	WEAK use in the rules
>0	FEW uses in the rules
=0	NO uses in the rules

Table 2-4 – Statistical Analysis used by Analyze Data

The *Analyze Values* tool was used to extract the high level description of the behavioral model developed in the previous section. The figure below shows the use of all fuzzy values in the behavioral model developed from the case study data.

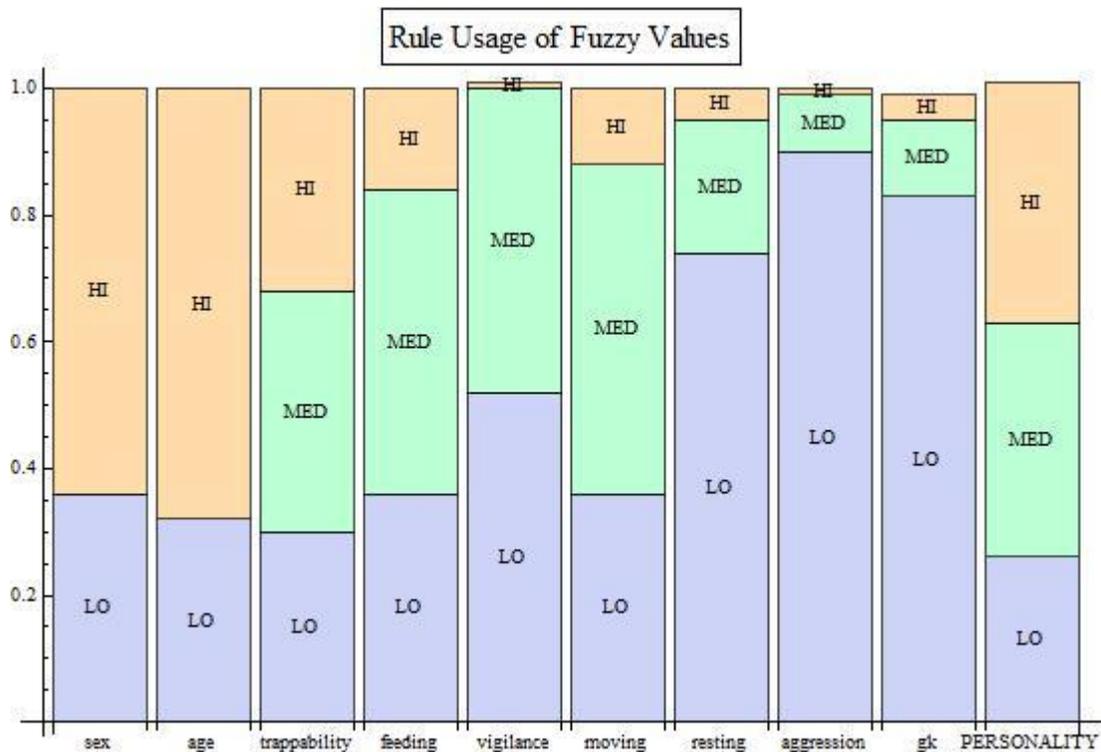


Figure 2-3 – Analyze Values output for All Rules

The bar chart above reflects the distribution of fuzzy values for all of the fuzzy variables found in the Rules file. Since the *Create Rules* tool was used to generate the rules from the data, the distribution of fuzzy values will not necessarily be equal but will instead reflect the occurrence of the fuzzy values in the data. Note that SEX and AGE reflect the use of only two values in the data. This is consistent with the data and the researcher's expectation. From the bar chart, we observe that fuzzy values used by the Consequent appear in nearly equal proportions. To better understand the impact of the fuzzy values on specific classification types, an analysis of the rules for each consequent fuzzy value was performed.

The bar chart below describes the use of fuzzy values in rule antecedents where the TIMID fuzzy value is used in the consequent of the rules. Observations regarding the TIMID Rules are detailed in the table below. Analysis of the TIMID rules provides some insight into the basic characteristics of TIMID Individuals.

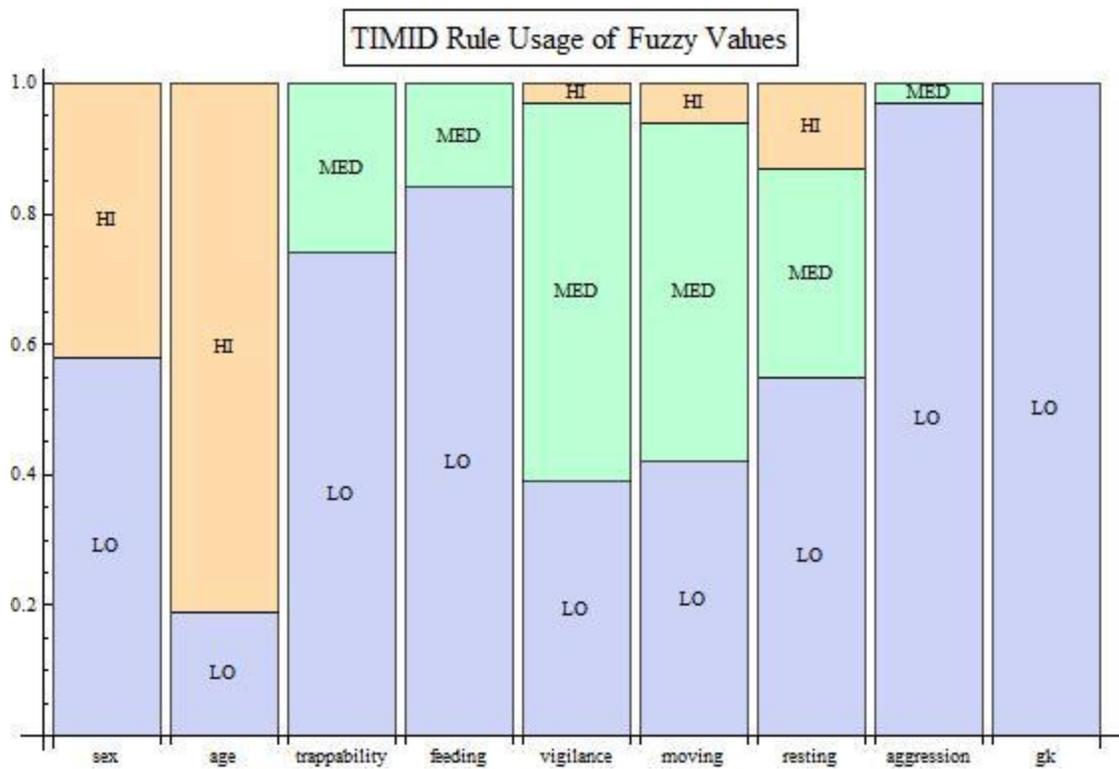


Figure 2-4 – Analyze Values output for TIMID Rules

Behavioral Model for TIMID Rules

sex has

- STRONG emphasis of M in the rules.
- SLIGHT emphasis of F in the rules.
- NO use of U in the rules.

age has

- STRONG emphasis of AD in the rules.
- WEAK use of YR in the rules.
- WEAK use of U in the rules.

trappability has

- STRONG emphasis of LOW in the rules.
- WEAK use of MED in the rules.
- NO use of HIGH in the rules.

feeding has

- STRONG emphasis of LOW in the rules.
- WEAK use of MED in the rules.
- WEAK use of HIGH in the rules.

vigilance has

- STRONG emphasis of MED in the rules.
- SLIGHT emphasis of LOW in the rules.
- FEW uses of HIGH in the rules.

moving has

- SLIGHT emphasis of MED in the rules.
- SLIGHT emphasis of LOW in the rules.
- FEW uses of HIGH in the rules.

resting has

- STRONG emphasis of LOW in the rules.
- AVERAGE use of MED in the rules.
- FEW uses of HIGH in the rules.

aggression has

- STRONG emphasis of LOW in the rules.
- WEAK use of MED in the rules.
- WEAK use of HIGH in the rules.

gk has

- STRONG emphasis of LOW in the rules.
- WEAK use of HIGH in the rules.
- WEAK use of MED in the rules.

Table 2-5 – Analyze Values output for TIMID Rules

The bar chart below describes the use of fuzzy values in rule antecedents where the NORMAL fuzzy value is used in the consequent of the rules. Observations regarding the NORMAL Rules are detailed in the table below.

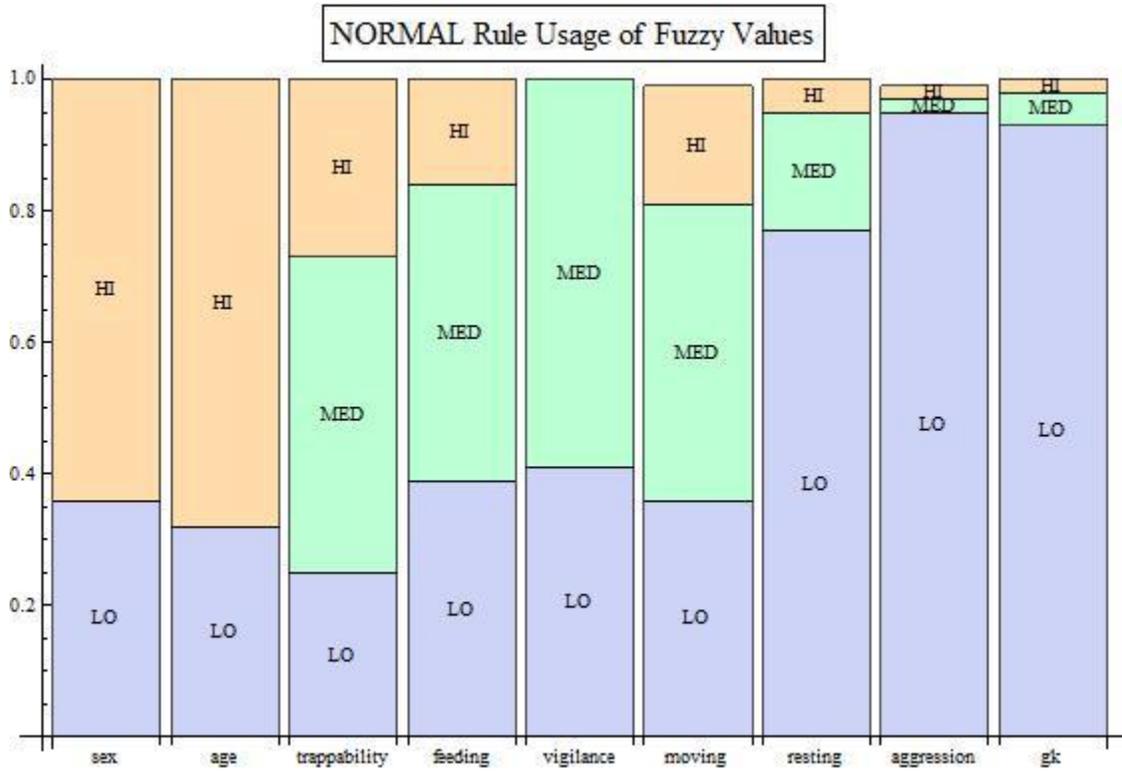


Figure 2-5 – Analyze Values output for NORMAL Rules

Behavioral Model for NORMAL Rules

sex has

- STRONG emphasis of F in the rules.
- SLIGHT emphasis of M in the rules.
- NO use of U in the rules.

age has

- STRONG emphasis of AD in the rules.
- AVERAGE use of YR in the rules.
- NO use of U in the rules.

trappability has

- STRONG emphasis of MED in the rules.
- WEAK use of HIGH in the rules.
- WEAK use of LOW in the rules.

feeding has

- SLIGHT emphasis of MED in the rules.
- SLIGHT emphasis of LOW in the rules.
- FEW uses of HIGH in the rules.

vigilance has

- STRONG emphasis of MED in the rules.
- SLIGHT emphasis of LOW in the rules.
- NO use of HIGH in the rules.

moving has

- STRONG emphasis of MED in the rules.
- SLIGHT emphasis of LOW in the rules.
- FEW uses of HIGH in the rules.

resting has

- STRONG emphasis of LOW in the rules.
- WEAK use of MED in the rules.
- WEAK use of HIGH in the rules.

aggression has

- STRONG emphasis of LOW in the rules.
- WEAK use of HIGH in the rules.
- WEAK use of MED in the rules.

gk has

- STRONG emphasis of LOW in the rules.
- WEAK use of MED in the rules.
- WEAK use of HIGH in the rules.

Table 2-6 – Analyze Values output for NORMAL Rules

The bar chart below describes the use of fuzzy values in rule antecedents where the BOLD fuzzy value is used in the consequent of the rules. Observations regarding the BOLD Rules are detailed in the table below.

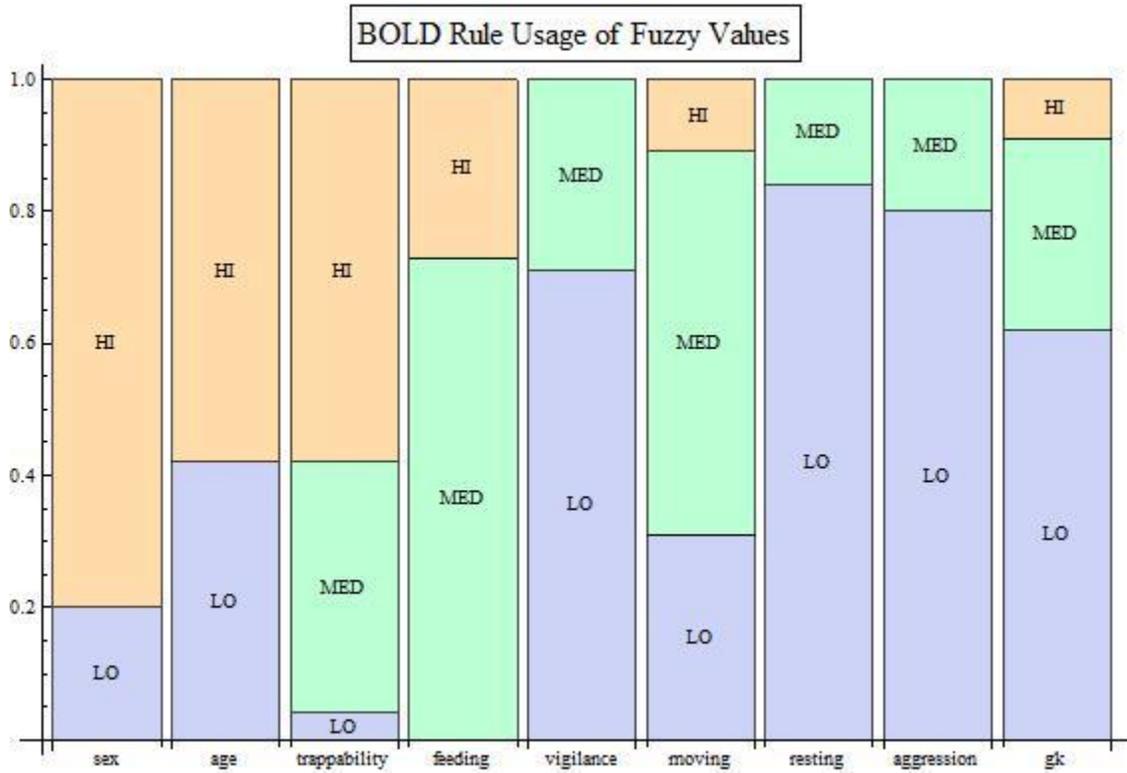


Figure 2-6 – Analyze Values output for BOLD Rules

Behavioral Model for BOLD Rules

sex has

- STRONG emphasis of F in the rules.
- WEAK use of M in the rules.
- NO use of U in the rules.

age has

- STRONG emphasis of AD in the rules.
- SLIGHT emphasis of YR in the rules.
- NO use of U in the rules.

trappability has

- STRONG emphasis of HIGH in the rules.
- SLIGHT emphasis of MED in the rules.
- FEW uses of LOW in the rules.

feeding has

- STRONG emphasis of MED in the rules.
- WEAK use of HIGH in the rules.
- NO use of LOW in the rules.

vigilance has

- STRONG emphasis of LOW in the rules.
- WEAK use of MED in the rules.
- NO use of HIGH in the rules.

moving has

- STRONG emphasis of MED in the rules.
- WEAK use of LOW in the rules.
- FEW uses of HIGH in the rules.

resting has

- STRONG emphasis of LOW in the rules.
- WEAK use of MED in the rules.
- WEAK use of HIGH in the rules.

aggression has

- STRONG emphasis of LOW in the rules.
- WEAK use of MED in the rules.
- WEAK use of HIGH in the rules.

gk has

- STRONG emphasis of LOW in the rules.
- WEAK use of MED in the rules.
- FEW uses of HIGH in the rules.

Table 2-7 – Analyze Values output for BOLD Rules

The goal of the *Analyze Values* tool is to provide a view of the behavioral model that facilitates the validation of the model by an expert in the problem domain. Viewing

the behavioral model from the perspective provided by the *Analyze Values* tool allows the expert to make high level statements about the classification types from the model. The table below contains a high level description of the classification types based on an interpretation of the output of the *Analyze Values* tool.

Classification Type	High level statement about the classification type
TIMID	Is difficult to trap, has heightened vigilance, does not eat as much as other classification types, rarely rests, is not aggressive and rarely greet kisses.
NORMAL	Has moderate behavior traits with the exception of not being aggressive and rarely greet kisses.
BOLD	Is not afraid of being trapped, is not as vigilant as others in the group, rests often, is not aggressive and rarely greet kisses

Table 2-8 – High level statements about the classification types

Being able to translate the data into high level statements allows the researcher to easily compare their knowledge of animal behavior with the observation data as classified by the BMA.

Organizing the data according to the classification types also allows the researcher to consider the behavioral model as a whole rather than a collection of rules. For example, the rules for TIMID, NORMAL and BOLD individuals all include a high proportion of rules with GK=LOW. The researcher may decide the membership function needs adjustment or that the variable is not providing a significant contribution to the model. In some cases, simplification of the behavioral model through the elimination of fuzzy variables from the fuzzy rules may improve the usefulness or accuracy of the model.

2.4 Classifying Individuals with the Behavioral Model

In addition to understanding the behavior of the group, the researcher in the case study used the classification index of the behavioral model to provide a measure for each individual represented in the data. The *Evaluate Data* tool of BMA is used to generate a classification index from the individual's data. The rule set developed in the previous section is sufficient to generate classifications for the data that was used to develop the rules. However, if new data becomes available, the rules may need to be expanded. Using the *Data* file, *Membership Function* file and the *Rule* file, the *Evaluate Data* tool generates a classification type and a classification index. A portion of the output of the *Evaluate Data* tool is presented in the table below.

Line	Classification Index	FuzzyValue	ClassificationType
2	9.9566	NORMAL	2
3	2.64	TIMID	1
4	17.2967	BOLD	4

Table 2-9 – Evaluate Rules output

3 Investigating relationships with the Behavioral Model: A Case Study

During the 1990's, Genetic Fuzzy Systems (GFS) based on Fuzzy Rule Based Systems (FRBS) coupled with a Genetic Algorithm (GA) were developed in response to the FRBS lack of learning ability [Cordón]. The BMA uses a GA to adjust the parameters of the behavioral model to investigate the occurrence of a relationship between the classification index and another measure of behavior.

Social Network Analysis has been used by biologists to study the communication between individuals in a group [Krause]. The primary focus of the research has been the mathematical analysis of the network representing individuals in a group. In this work, the researcher was interested in determining if there was a relationship between the Social Network Analysis metrics of the individuals and the classification index of the individuals obtained from the behavioral model.

To determine if a relationship exists between the Social Network Analysis metrics and the classification index is a straightforward statistical problem. However, to fully investigate if any relationship exists between the classification indexes from the FRBS and Social Network Analysis metrics, alternative values for the modal points of the membership functions should be explored. BMA adjusts the membership functions of the behavioral model to improve the relationship between the classification index and the Social Network Analysis metric. Using expert knowledge, the researcher evaluates the new membership functions to determine if an accurate representation of the activity being modeled has been obtained.

Since little is known about the type of relationship that might exist between the classification index and the Social Network Analysis metric, the GA fitness function uses a statistical analysis technique to determine if a monotonically increasing or decreasing relationship can be found. The following sections describe specifically how the GA is used to adjust the parameters of the model and the results of the case study.

3.1 Chromosome Encoding

BMA uses Pittsburgh style encoding of the chromosome where the complete Data Base (DB) of the FRBS is encoded in a chromosome. The GA encodes a fuzzy variable as a gene. Associated with each fuzzy variable are three fuzzy values. The alleles of the gene are the membership functions for the fuzzy values associated with the fuzzy variable. Consequently, the chromosome contains genes representing each of the membership functions of the FRBS. BMA provides a mechanism for excluding the genes from the evolutionary process. The researcher may choose to exclude genes representing well understood variables in the behavioral model which do not require further investigation or variables with known values such as SEX.

Due to the encoding of the membership functions in the alleles, changes to the genes during initialization, reproduction, and mutation are constrained. The design of the membership functions associated with a fuzzy variable constrains the modifications that may be performed on the value of the genes. For example, given a fuzzy variable with three fuzzy values, LOW, MEDIUM and HIGH, the crisp sets of the membership functions describing the three fuzzy values should maintain the linguistic representation. In other words, the crisp values for LOW should not be greater than the crisp values of MEDIUM. Also of concern is the overlap of two membership functions. While overlap

is acceptable and provides for partial membership, two membership functions should not overlap to the point that each membership function could have full membership for a crisp value.

During the initialization of the population of chromosomes, the maximum and minimum crisp values of the fuzzy variable are used as limits on the values of the genes representing the fuzzy variable. A random value between the maximum and minimum values is selected as the midpoint of the MEDIUM membership function. A second random value between the maximum and minimum values is selected to define the width of the base of the trapezoid. A third random value is selected as a multiplier to scale the base of the trapezoid and define the top of the trapezoid. The figure below illustrates the construction of the first of the three membership functions for a fuzzy variable.

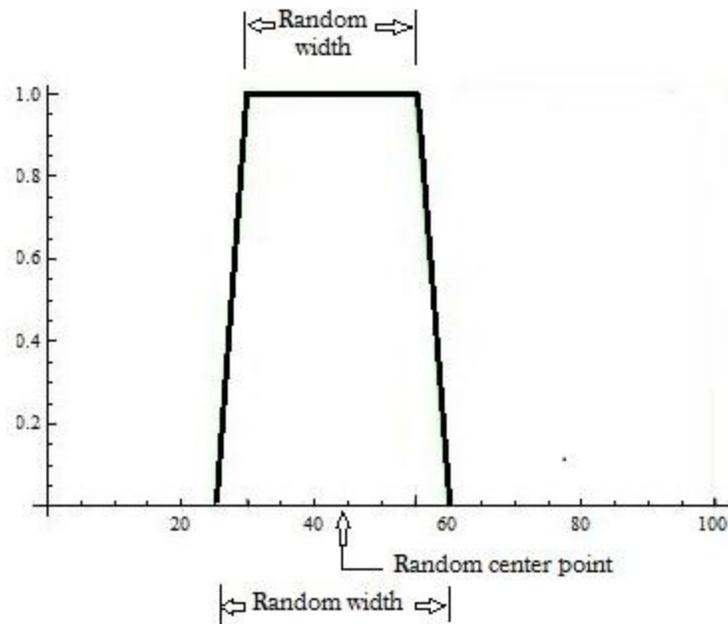


Figure 3-1 – Initial random membership function definition

When randomizing the gene, performing crossover or mutation, the proposed values must be evaluated to ensure the trapezoid formed is a legal membership function. The random nature of the center position, base width and top width must conform to acceptable trapezoid dimension in order to be used as a membership function. The table below describes the tests placed on the randomly defined or modified membership function.

Defect	Corrective Action
Values out of sequence	Sort to ascending values
All points have the same value	Randomize
Base wider than Max Range	Scale to fit Max Range
Square	Scale the top by ratio of base to Max width
Below Minimum Value	Translate to Minimum value
Above Maximum Value	Translate to Maximum value
Left Triangle or Trapezoid	Convert to triangle
Right Triangle or Trapezoid	Convert to triangle

Table 3-1 – Constraint test for valid trapezoid

Randomly defined or modified membership functions which are found to be defective are corrected to ensure a valid gene is created and avoid premature convergence due to the creation of poor solution candidates.

The MEDIUM membership function is then used to define LOW and HIGH membership functions. The LOW and HIGH membership functions are either trapezoids or triangles depending on the randomized values of the MEDIUM membership function. For the LOW membership function, the first two values are taken from the Minimum Value for the Fuzzy Variable. The third and fourth values are taken from the first two values of the middle membership function. For the HIGH membership function, the first two values are taken from the last two values of the MEDIUM membership function. The third and fourth values are taken from the Maximum Value of the fuzzy variable. The assignment algorithm ensures a valid relationship between the membership functions for a fuzzy variable. The figure below shows the addition of the first and third membership functions to the initial randomly defined membership function encoded as a gene.

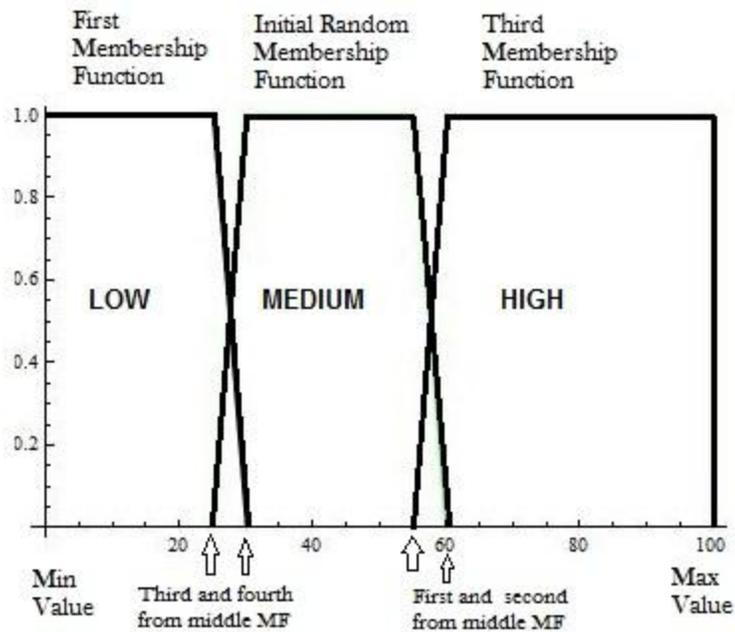


Figure 3-2 – Initial randomly defined gene

3.2 Fitness Function

The fitness function is used by the GA to determine the value of the chromosome. The value of the chromosome is used in the selection of chromosomes during the Reproduction and Natural Selection phases of the evolutionary process. Higher valued, or more fit, chromosomes indicate the set of genes encoded in the chromosome have provided a better solution to the problem being solved.

The investigation of the relationship between the classification index and another measure of behavior implemented by the fitness function imposes the following requirements on the fitness function:

- Individual classification type, that is, the fuzzy value of the individual, is maintained through the evaluation. Changing the membership functions will change the fuzzy values represented by the crisp inputs, but changes to the membership functions should not change the classification type of the individual. For our case study, a TIMID individual cannot become BOLD.
- The form of the relationship is not restricted to well-known functions such as linear, exponential, etc.
- The classification index is not required to exactly match the external measure.

A monotonic relationship is one of the most basic mathematical relationships between two variables in data sets. Monotonic functions tend to increase or decrease in only one direction.

Spearman's rank correlation coefficient is a statistical measure of the covariance of two variables. The Spearman correlation coefficient ranges between -1 and +1 with the extremes denoting the rank of the two variables can be described by a perfect monotonic function. A monotonically increasing function of two variables, x and y , where $x \leq y$, is a function where $f(x) \leq f(y)$. If the order is reversed such that $f(x) \geq f(y)$, the function is known as monotonically decreasing. The Spearman correlation coefficient is a measure of the rank of two variables and is concerned with the order of the variables rather than the values of the variables. As such, the Spearman Correlation Coefficient is an excellent measure of the relationship between two variables of different scale types where order is maintained or two variables of the

same scale type where the range of the variables is different. The Spearman Correlation Coefficient provides a good measure of the fitness of the chromosome by proving a monotonic relationship exists between the classification index and another measure of behavior.

The Spearman Correlation Coefficient is defined as the Pearson Correlation Coefficient of the ranked variables. The Spearman Correlation Coefficient is expressed by the equation:

$$rho = \frac{\sum i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum i (x_i - \bar{x})^2 \sum i (y_i - \bar{y})^2}}$$

Where x_i and y_i are the rank of the two variables and \bar{x} and \bar{y} are the average of the ranks of the two variables.

The null hypothesis under investigation is the Spearman Coefficient is obtained by a random ordering of the two variables, in other words, that no relationship exists between the classification index and another measure of behavior. The evaluation of the null hypothesis is performed based on the P value calculated from the Spearman Correlation Coefficient. The P value is the probability of getting the observed result if the null hypothesis is true. A low P value implies the null hypothesis can be rejected. In our case, a low P value indicates the two variables, the classification index and another measure of behavior, do covary to the degree expressed.

Adjusting the membership functions can potentially change the fuzzy value that represents a measure of activity. In other words, the measure of an activity, such as FEEDING, may be either LOW or MEDIUM. However, the behavioral model

classification of the individual should not change – a BOLD individual should not become TIMID. Two different methods, Strict and Loose, are provided in the fitness function to maintain the classification type during the GA process. BMA may be configured to use the method best suited for a specific investigation. Both methods improve the performance of the fitness function by restricting the rule set used by the fitness function. For both methods, the individuals are evaluated using the initial membership functions to establish a baseline classification type. The resulting classification types should be verified by an expert prior to investigating the relationship between the classification index and another measure of behavior. The Strict evaluation method maintains the classification type by limiting the rule set used by the fitness function to the set of rules that were active during the initial evaluation of the classification type. The Loose evaluation method maintains classification type but uses a larger set of rules. The Loose evaluation method limits the rules used by the fitness function to the set of rules that conform to the classification type. Loose evaluation uses a superset of the rules found by the Strict method.

Seeking the relationship between the classification index and another measure of behavior by modifying the membership functions impacts the chance of finding an active rule for an individual regardless of which evaluation method is selected. Recall that the observation data is fuzzified and the fuzzy inference used to generate a fuzzy output based on the discovery of a rule where the fuzzy implication operator on the antecedents finds the rule is at least partially true. As the membership values are changed from chromosome to chromosome over generations, the fuzzified values of the observational data will change. While a rule may exist that maintains the individual's

classification type, there is also a chance that the new fuzzy values no longer activate a rule of the correct classification type.

If no rule is active for the individual, the individual should not be included in the Spearman coefficient analysis of the chromosome. However, simply excluding individuals without active rules during fitness evaluation is not optimum since the Spearman Correlation Coefficient for the chromosome is artificially inflated. To compensate for individuals without active rules, the BMA penalizes the chromosome proportionally to the number of individuals without active rules. The fitness function for the chromosome is then the combination of the following steps:

- Evaluate each individual using either STRICT or LOOSE method recording the number of individuals with active rules.
- Calculate the Spearman Correlation Coefficient between the evaluation and the selected measure of behavior.
- Calculate the fitness value as the product of the Spearman Correlation Coefficient for individuals with active rules and the ratio of individuals with active rules to the total number of individuals in the data set.

3.3 Reproduction, Mutation and Natural Selection

During the Reproduction process, chromosomes from the parent population are combined to create the next generation population of chromosomes. The crossover operator and the method of choosing the next generation population from the parent and children populations were chosen with the goals of maintaining diversity in the population and avoiding premature convergence.

The choice to encode the fuzzy variables as genes coupled with the choice of a crossover operation at the gene level produces new chromosomes that conform to the requirements imposed by the membership functions. The crossover operator implemented by the GA follows the classic strategy at the gene level rather than the allele level. The crossover operator is applied to two randomly selected chromosomes. A random number known as the crossover point is used to select a point in the gene sequence. The genes of one chromosome before the crossover point are combined with the genes from the second chromosome after the crossover point as shown in the figure below.

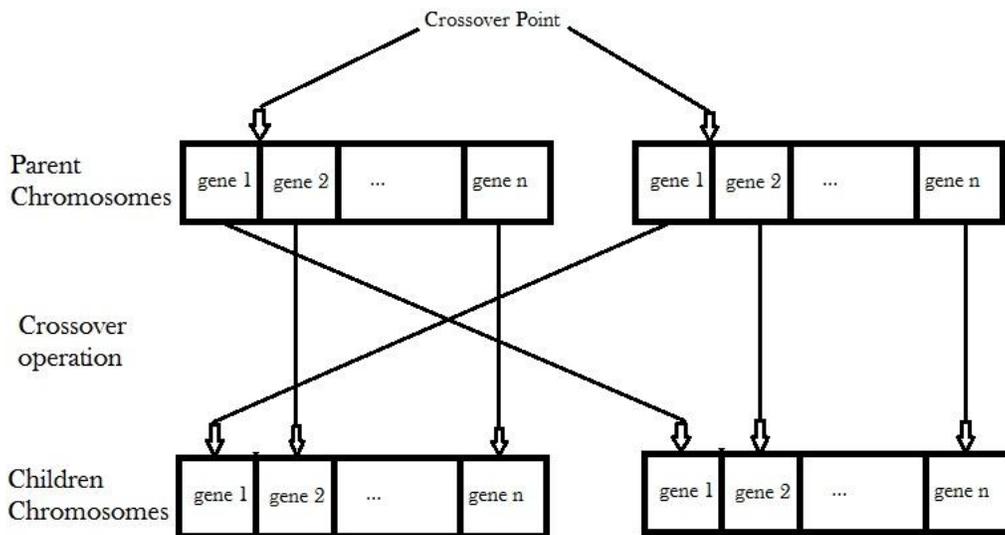


Figure 3-3 - Gene level crossover

The selection of the mating pairs from the population of chromosomes must avoid the repeated use of the same chromosome to avoid a decrease in the diversity of the population. For each chromosome, the GA randomly selects another chromosome from the population as a mate. Chromosomes are not allowed to mate with themselves or with another chromosome more than once. The probability of crossover for each chromosome is 100%. The selection of the population as the result of reproduction uses the fitness of each chromosome in both the parent and children populations to select the most fit chromosomes for the next generation. This algorithm ensures that good solutions from the previous generation are not lost if a better solution has not been found through the crossover operation. Mutation is used to ensure diversity in the new population by making small changes in a single chromosome. Mutation like crossover is performed by the BMA at the gene level. A chromosome is randomly selected from the candidate population produced by the reproduction process. A gene within the selected chromosome is randomly selected. Finally, the values of the gene are randomized in the same manner described during the initialization of the chromosome. All other genes encoded in the chromosome remain unchanged. The probability of mutation of a chromosome within the population is 0.1%.

Having created a new population through the reproduction and mutation phases, the next generation population must be selected from the candidate population. The natural selection phase strives to create a population with the best possible chromosomes while avoiding premature convergence on a local optimum. BMA ranks the candidate population based on the fitness value of each chromosome. If more than one chromosome exists for each rank, BMA selects the most fit individual from each

rank until all of the ranks have been evaluated. BMA then repeats the process selecting the next most fit individual from each rank until next generation population has been filled. This strategy improves diversity by minimizing the selection of individuals from the same rank ensuring a mix of high and low fitness chromosomes.

3.4 Application to the Case Study

As previously described, a behavioral model was extracted from the data obtained from observing a group of individual's activities. The behavioral model was used to classify the individuals. The behavioral model used nine fuzzy input variables each with three possible fuzzy values. To completely describe the behavioral model would require 19,683 rules to cover every possible combination of fuzzy value and fuzzy variable. The 120 rules extracted from the data reflect a small percentage of the possible rules. Initial tests showed that maintaining the behavioral model classification type and obtaining an acceptable percentage of individuals with active rules while adjusting the membership functions on such a small behavioral model was unlikely.

After evaluating the behavioral model developed in the section above, the researcher decided to remove two variables from the equation to reduce the size of the rule set. To improve the ability of BMA to find a relation, the *Rule* file was expanded to include every possible combination of fuzzy values in a rule. The *Generate Rules* tool was used to build a set of proposed rules that covered all 972 possible rules based on the use of the fuzzy values in the data. The researcher used their expert knowledge to assign the consequent values for the new *Rule* file.

The Loose fitness evaluation was selected after testing showed an improvement in the ability of BMA to find a set of membership functions which resulted in active rules for a greater number of individuals. The table below compares average results of tests using the Loose and Strict fitness evaluation methods.

Fitness Evaluation	Individuals with active rules
Strict	69%
Loose	75%

Table 3-2 – Comparison of Fitness Evaluation Methods

The next consideration in the design of the tests was to determine the best size for the population. Test cycles were performed with increasing population size. Due to the initialization of the chromosomes and the probabilities of mutation and crossover, the GA will not consistently choose the same solution and the time to convergence will vary for the same population sizes. Samples were taken and averaged to discover a population size that balanced the trade-off between time to convergence, the percentage of individuals with active rules and the Spearman Coefficient. The tests showed that for populations greater than 2,000 chromosomes, little increase in Spearman coefficient or the percentage of individuals with active rules was achieved despite a significant increase in the time to convergence. Taking into account the three objectives, a population size of 2,000 was selected for the remaining tests. The table below shows a portion of the test results.

Population Size	Spearman	P Value	Active Individuals	Average Time To Convergence
100	0.5581	0.016	55%	0.042 hours
1,000	0.6514	0.003	61%	0.38 hours
2,000	0.6749	0.002	62%	0.95 hours
5,000	0.6581	0.003	62%	2 hours
10,000	0.7039	0.001	62%	4 hours

Table 3-3 – Comparison of population size test results

The researcher had four Social Network Analysis metrics to investigate. The Social Network Analysis metrics were Betweenness, Degrees, Path Length and Total Interactions. Betweenness is a measure of a node’s centrality in the network. Degree is a measure of the relationship between other nodes in the network. Path Length is a measure of distance between nodes in the network. Total Interactions measures the communication between a node and other nodes in the network. The *Correlation* tool was used to investigate the relationship between classification indices and the four Social Network Analysis metrics.

To perform the tests, the *Correlation* tool was configured to use the following settings:

- a fitness function based on Spearman Correlation Coefficient and the Loose GA Evaluation method with penalty imposed proportional to the percentage of individuals with active rules.
- population size of 2,000.
- Exclusion of AGE, SEX and PERSONALITY fuzzy variables from the GA membership function adjustment process.

The GA adjusts the membership function values seeking a set of membership functions that result in classification indices related to the Social Network metric being investigated. The final set of membership functions produced by the GA must be evaluated by the researcher to confirm the crisp values provide reasonable representation of the fuzzy values over the universe of discourse.

SEX, AGE and PERSONALITY were excluded from the GA membership function adjustment process. The SEX and AGE data represent known values of the individuals. PERSONALITY is the fuzzy variable used in the consequent of the rules. Adjusting the membership functions associated with PERSONALITY would change the classification index but have no effect on the classification type which is maintained throughout the GA processing.

3.4.1 Complete Data Set Tests

The researcher provided a *Data* file containing the observations of 80 individuals and four Social Network Analysis metrics for each of the individuals. The *Data* file was evaluated using the *Membership Functions* file defined by the researcher to establish a baseline classification type for each individual. The BMA *Correlation* tool was executed to determine if a set of membership functions could be found with a higher degree of correlation than the original set of membership functions. The test results are shown in the table below.

Social Network Metric	Initial Evaluation			Correlation Test Results		
	Spearman	P-Value	Active	Spearman	P-Value	Active
Betweenness	0.238	0.034	100%	0.6053	0.000003	63%
Degrees	0.229	0.041	100%	0.5297	0.000008	79%
Path Length	-0.03	0.793	100%	-0.5422	0.0004	49%
Total Interactions	0.383	0.0004	100%	0.5698	0.0000001	91%
Averages	0.22		100%	0.56175		71%

Table 3-4 – Correlation Test Results

From the test results, we observe BMA was able to find a set of membership functions for each of the Social Network Analysis metrics that improved the probability that a relationship exists between the classification index produced by the FRBS and the Social Network Analysis metric. In each test case, we observe that the percentage of individuals with Active Rules is less than 100% indicating that BMA was unable to find a set of membership functions that maintained the classification index of every individual. The modified membership functions would have caused reclassification of the individuals for 29% of the individuals on average. Individuals that did not have active rules were not considered in the Spearman Correlation Coefficient calculation.

For three of the Social Network Analysis metrics, Betweenness, Degrees, and Total Interactions, BMA discovered a monotonically increasing relationship between the classification index and the Social Network Analysis metrics. A monotonically increasing relationship indicates that TIMID individuals will have lower values of these Social Network Analysis metrics than BOLD individuals. For the Path Length, BMA discovered a monotonically decreasing relationship indicating that TIMID individuals will have higher values of the Path Length metric than BOLD individuals.

Comparing the membership functions produced by the Correlation Tests to the original membership functions allows the researcher to investigate factors influencing the relationship between the classification index and the Social Network Analysis. The adjustment to the TRAPPABILITY membership function during the Correlation Test for Total Interactions is shown in the figure below. The blue trapezoids indicate the original set of membership functions for TRAPPABILITY and the green trapezoids indicate the adjusted set of membership functions.

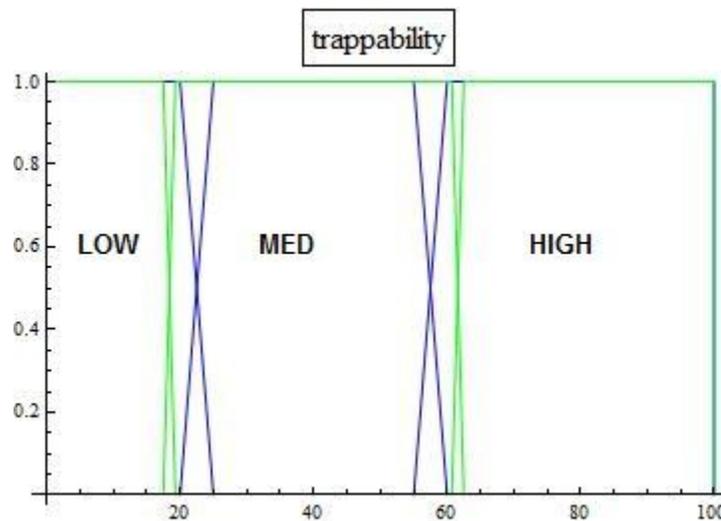


Figure 3-4 – Trappability Membership Function Comparison

From the figure above, we observe that the modal points defining MED for TRAPPABILITY were increased to improve the correlation between the classification index and the Total Interactions Social Network Analysis metric. The original membership functions were based on the researcher's expert knowledge. The adjusted membership functions are the result of considering the values required by the behavioral model to improve the relationship between the behavioral model classification index and the Total Interactions metric. The adjusted membership

function values provide useful guidelines for investigating the impact of variables and their values on the behavioral model.

3.4.2 Group Level Tests

Contained in the *Data* file were three different sets of observational data representing individuals grouped in colonies. An individual was only present in one of the colony data sets. The test described above analyzed all of the colonies as one data set. To further understand the relationships of individuals within a colony, the Correlation Tests were repeated on the groups, or colonies, of individuals contained in the *Data* file.

Test Group	Initial Evaluation			Correlation Test Results		
	Spearman	P-Value	Active	Spearman	P-Value	Active
1	0.076	0.6992	100%	0.7039	0.001	62%
2	0.67	0.0003	100%	0.8202	0.000001	96%
3	0.348	0.0812	100%	0.6462	0.001	85%
Averages	0.364667		100%	0.723433		81%

Table 3-5 – Betweenness Test Results

The test results indicate BMA is able to find a set of membership functions that support a higher degree of correlation between the groups than the original membership functions. As can also be observed from the test results for the percentage of active individuals, BMA was not able to find a set of membership functions that maintained the classification type for all of the individuals in a group. The modified membership functions would have caused reclassification of the individuals for 19% of the individuals on average. Individuals that did not have active rules were not considered in the Spearman Correlation Coefficient. From the *Correlation Test Report*, the researcher

inspects the results to draw conclusions about their meaning in the context of the researcher's investigation into animal behavior. The test results for the remaining Social Network Analysis metrics are shown in the tables below.

Test Group	Initial Evaluation			Correlation Test Results		
	Spearman	P-Value	Active	Spearman	P-Value	Active
1	0.075	0.706	100%	0.7721	0.0002	62%
2	0.466	0.019	100%	0.7623	0.00006	84%
3	0.389	0.0496	100%	0.713	0.0004	78%
Averages	0.31		100%	0.749133		75%

Table 3-6 – Degrees Test Results

Group	Initial Evaluation			Correlation Test Results		
	Spearman	P-Value	Active	Spearman	P-Value	Active
1	-0.248	0.1942	100%	-0.8712	0.000003	72%
2	-0.002	0.9931	100%	-0.6772	0.0008	56%
3	0.111	0.5908	100%	0.8373	0.0009	50%
Averages	-0.04633		100%	-0.23703		59%

Table 3-7 – Path Length Test Results

Group	Initial Evaluation			Correlation Test Results		
	Spearman	P-Value	Active	Spearman	P-Value	Active
1	0.271	0.1557	100%	0.7789	0.00009	66%
2	0.593	0.0018	100%	0.7643	0.00001	96%
3	0.381	0.0547	100%	0.6311	0.0012	89%
Averages	0.415		100%	0.724767		84%

Table 3-8 – Total Interactions Test Results

4 Conclusions

The use of the Behavioral Model Analyzer (BMA) in conjunction with the protocol to build a behavioral model was demonstrated in the case. The behavioral model integrates the analysis of data obtained from the observation of the activities of the individual in a social context with Social Network Analysis and provides a quantitative basis for assessing the impact of activities on an individual's classification. By providing a behavioral model consisting of a well-defined structure and set of parameters, each element of the model can be further investigated to discover knowledge about the activities which effect behavior.

In the case study, a high level, descriptive model of the behavior was extracted from the observational data. The descriptive behavioral model allowed the researcher to easily compare their expert knowledge with the fuzzy rules used to construct the model. By providing a high level linguistic model, the study showed that information about a large data set may be readily accessed by researchers without requiring a detailed understanding of fuzzy set theory. The behavioral model developed was used to generate a classification type and classification index for each individual in the observational data. In the case study, BMA was used to explore the relationship between the classification indexes and Social Network Analysis metrics. The resulting membership functions are useful in investigating the relationship to the individuals, their classification and the Social Network Analysis metrics.

While the development of the behavioral model using the tool and the protocol produced a descriptive model, the investigation into the relationship between the

classification index produced by the behavioral model and the Social Analysis metric highlights several issues that could be explored in future work. The defuzzification of the fuzzy inference was of limited use as a classification index. Many of the individuals in the group had only one active rule during evaluation resulting in the same classification index for each individual of that classification type. Consequently, the crisp output of the defuzzification has limited usefulness in providing a measure beyond a basic classification value in the cases of individuals with only one active rule of the same classification type. Further investigation into defuzzification methods that provide a weight to the crisp output for individuals with the same fuzzy output value would improve the usefulness of BMA in this area.

The design of the fitness function is critical to the operation of the GA and the case study highlights the difficulty of maintaining the classification types of the individuals as the membership functions are adjusted. Recent research concerning the use of GFS to solve decision making problems is focused on the use of multi-objective fitness functions. A good decision is based on knowing the appropriate balance of multiple goals. Multi-objective fitness functions strive to find the best solution from several measures of fitness such as the Spearman Correlation Coefficient, P value and the percentage of individuals with active rules. While the BMA addressed multiple goals with the introduction of a penalty for individuals without active rules, more complex measures such as seeking a Pareto Optimum solution [Sánchez] may improve the results.

REFERENCES

- Bakirli, Gozde; Birant, Derya; Kut, Alp (March 2011) An incremental genetic algorithm for classification and sensitivity analysis of its parameters, *Expert Systems with Applications*, Volume 38, Issue 3, March 2011, Pages 2609-2620, ISSN 0957-4174, DOI: 10.1016/j.eswa.2010.08.051.
- Berkan, Riza C. and Trubatch, Sheldon L. (1997) Fuzzy Systems Design Principles: Building Fuzzy IF-THEN Rule Bases, *IEEE Press*.
- Chang, Pei-Chann ; Huang, Wei-Hsiu; Ting, Ching-Jung (2010) Dynamic diversity control in genetic algorithm for mining unsearched solution space in TSP problems, *Expert Systems with Applications*, Volume 37, Issue 3, 15 March 2010, Pages 1863-1878, ISSN 0957-4174, 10.1016/j.eswa.2009.07.066.
- Cordón, Oscar (September 2011) A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems, *International Journal of Approximate Reasoning*, Volume 52, Issue 6, September 2011, Pages 894-913, ISSN 0888-613X, 10.1016/j.ijar.2011.03.004.
- Cordón, O; Gomide, F.; Herrera, F.; Hoffmann, F.; Magdalena, L. (2004) Ten years of genetic fuzzy systems: current framework and new trends, *Fuzzy Sets and Systems* 141 (1) (2004) 5–31.

Cowgill, M. C.; Harvey, R. J.; Watson, L. T. (April 1999) A genetic algorithm approach to cluster analysis, *Computers & Mathematics with Applications*, Volume 37, Issue 7, April 1999, Pages 99-108, ISSN 0898-1221, DOI: 10.1016/S0898-1221(99)00090-5.

Deb, Kalyanmoy ; (2000) An efficient constraint handling method for genetic algorithms, *Computer Methods in Applied Mechanics and Engineering*, Volume 186, Issues 2–4, 9 June 2000, Pages 311-338, ISSN 0045-7825, 10.1016/S0045-7825(99)00389-8.

Deng, P. S.; Tsacle, E. G. (October 2000) Coupling genetic algorithms and rule-based systems for complex decisions, *Expert Systems with Applications*, Volume 19, Issue 3, October 2000, Pages 209-218, ISSN 0957-4174, DOI: 10.1016/S0957-4174(00)00033-6.

Huang, H.; Pasquier, M.; Quek, C. (September 2011) Decision support system based on hierarchical co-evolutionary fuzzy approach: A case study in detecting gamma ray signals, *Expert Systems with Applications*, Volume 38, Issue 9, Pages 10719-10729.

Herrera, F.; Lozano, M.; Verdegay, J. L. (November 1998) A learning process for fuzzy control rules using genetic algorithms, *Fuzzy Sets and Systems*, Volume 100, Issues 1-3, 16 November 1998, Pages 143-158, ISSN 0165-0114, DOI: 10.1016/S0165-0114(97)00043-2.

Holland, J. H. (1974) *Adaptation in Natural and Artificial Systems*, *University Of Michigan Press*

Holland, J. H. (1992) Genetic Algorithms, *Scientific America*, Vol. 267, pp. 66-72

Hummon, Norman P. ; Doreian, Patrick (December 1990) Computational methods for social network analysis, *Social Networks*, Volume 12, Issue 4, December 1990, Pages 273-288, ISSN 0378-8733, DOI: 10.1016/0378-8733(90)90011-W.

James, Richard; Croft, Darren P.; Krause, Jens

Krause, Jens; Lusseau, David; James, Richard (2009 May) Animal social networks: an introduction, *Behavioral Ecology and Sociobiology*, Vol. 63.7 May 2009, Pages 967-973.

Kim, Myoung-Jong; Han, Ingoo (November 2003) The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms, *Expert Systems with Applications*, Volume 25, Issue 4, November 2003, Pages 637-646, ISSN 0957-4174, DOI: 10.1016/S0957-4174(03)00102-7.

Konak, Abdullah ; Coit, David W.; Smith, Alice E. (2006) Multi-objective optimization using genetic algorithms: A tutorial, *Reliability Engineering & System Safety*, Volume 91, Issue 9, September 2006, Pages 992-1007, ISSN 0951-8320, 10.1016/j.ress.2005.11.018.

Kwong, C.K.; Chan, K.Y.; Tsim, Y.C. (March 2009) A genetic algorithm based knowledge discovery system for the design of fluid dispensing processes for electronic packaging, *Expert Systems with Applications*, Volume 36, Issue 2, Part 2, March 2009, Pages 3829-3838, ISSN 0957-4174, DOI: 10.1016/j.eswa.2008.02.041.

- Liang, Yong; Leung, Kwong-Sak (2011) Genetic Algorithm with adaptive elitist-population strategies for multimodal function optimization, *Applied Soft Computing*, Volume 11, Issue 2, March 2011, Pages 2017-2034, ISSN 1568-4946, 10.1016/j.asoc.2010.06.017.
- Liao, Shu-Hsien (January 2005) Expert system methodologies and applications--a decade review from 1995 to 2004, *Expert Systems with Applications*, Volume 28, Issue 1, January 2005, Pages 93-103, ISSN 0957-4174, DOI: 10.1016/j.eswa.2004.08.003.
- Lusseau D; Newman MEJ (2004) Identifying the role that animals play in their social networks. *Proceedings of the Royal Society of London Series B-Biological Sciences* 271:S477-S481.
- Mamdani, E.H. and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, 1975.
- Marijtje A.J. van Duijn, Jeroen K. Vermunt, What Is Special About Social Network Analysis?, *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, Volume 2, Issue 1, March 2006, Pages 2-6, ISSN 1614-1881, DOI: 10.1027/1614-1881.2.1.2.
- McDonald, John H. (2009) *Handbook of Biological Statistics, Second Edition, Sparky House Publishing.*

- Mirhosseyeni, S. Hamid L.; Webb, Phil (November 2009) A Hybrid Fuzzy Knowledge-Based Expert System and Genetic Algorithm for efficient selection and assignment of Material Handling Equipment, *Expert Systems with Applications*, Volume 36, Issue 9, November 2009, Pages 11875-11887, ISSN 0957-4174, DOI: 10.1016/j.eswa.2009.04.014.
- Quagliarella, D. (1998) Genetic Algorithms and Evolution Strategy in Engineering and Computer Science: Recent Advances and Industrial Applications, *John Wiley & Sons, Ltd.*
- Romero, Txomin; Larrañaga, Pedro (2009) Triangulation of Bayesian networks with recursive estimation of distribution algorithms, *International Journal of Approximate Reasoning*, Volume 50, Issue 3, March 2009, Pages 472-484, ISSN 0888-613X, 10.1016/j.ijar.2008.09.002.
- Roy, Shibendu Shekhar (April 2006) Design of genetic-fuzzy expert system for predicting surface finish in ultra-precision diamond turning of metal matrix composite, *Journal of Materials Processing Technology*, Volume 173, Issue 3, 20 April 2006, Pages 337-344, ISSN 0924-0136, DOI: 10.1016/j.jmatprotec.2005.12.003.
- Sánchez, Luciano; Couso, Inés ; Casillas, Jorge (2009) Genetic learning of fuzzy rules based on low quality data, *Fuzzy Sets and Systems*, Volume 160, Issue 17, 1 September 2009, Pages 2524-2552, ISSN 0165-0114, 10.1016/j.fss.2009.03.004.

- Sánchez, Luciano; Suárez, M. Rosario; Villar, J.R.; Couso, Inés (2008) Mutual Information-based feature selection and partition design in fuzzy rule-based classifiers from vague data, *International Journal of Approximate Reasoning*, Volume 49, Issue 3, November 2008, Pages 607-622, ISSN 0888-613X, 10.1016/j.ijar.2008.06.005.
- Seidman, Stephen B. ; Foster, Brian L. (1979-1980) Sonet-1: Social network analysis and modeling system, *Social Networks*, Volume 2, Issue 1, 1979-1980, Pages 85-90, ISSN 0378-8733, DOI: 10.1016/0378-8733(79)90013-3.
- Sih A, Bell AM, Johnson JC, & Ziemba RE (2004) Behavioral syndromes: An integrative overview. *Quarterly Review of Biology* 79(3):241-277.
- Sih A & Bell AM (2008) Insights for behavioral ecology from behavioral syndromes. *Advances in the Study of Behavior*, Vol 38, Advances in the Study of Behavior), Vol 38, pp 227-281.
- Sih A, Hanser SF, & McHugh KA (2009) Social network theory: new insights and issues for behavioral ecologists. *Behavioral Ecology and Sociobiology* 63(7):975-988.
- Subelj, Lovro; Furlan, Stefan; Bajec, Marko (January 2011) An expert system for detecting automobile insurance fraud using social network analysis, *Expert Systems with Applications*, Volume 38, Issue 1, January 2011, Pages 1039-1052, ISSN 0957-4174, DOI: 10.1016/j.eswa.2010.07.143.
- Tsoukalas, Lefteri H. and Uhrig, Robert E. (1997) Fuzzy and Neural Approaches in Engineering, *John Wiley & Sons, Inc.*

Verdolin, J.L. (2007) Resources, not male mating strategies, determine social structure in Gunnison's prairie dogs. *Behaviour* 144, 1361-1382.

Wang, Jyun-Cheng; Chiu, Chui-Chen (April 2008) Recommending trusted online auction sellers using social network analysis, *Expert Systems with Applications*, Volume 34, Issue 3, April 2008, Pages 1666-1679, ISSN 0957-4174, DOI: 10.1016/j.eswa.2007.01.045.

Wang, N.F.; Tai, K. (2010 October) Target matching problems and an adaptive constraint strategy for multiobjective design optimization using genetic algorithms, *Computers & Structures* Volume 88, Issues 19-20 pp. 1064-1076

Wang, Ying-Ming (2009) Centroid defuzzification and the maximizing set and minimizing set ranking based on alpha level sets, *Computers & Industrial Engineering*, Volume 57, Issue 1, August 2009, Pages 228-236, ISSN 0360-8352, 10.1016/j.cie.2008.11.014.

Wasserman, Todd (2000) Of diapers and beer, *Brandweek* 41.9 February 8, 2000, Page 34

Wilson ADM & Stevens ED (2005) Consistency in context-specific measures of shyness and boldness in rainbow trout, *Oncorhynchus mykiss*. *Ethology* 111(9):849-862.

Wilson DS, Coleman K, Clark AB, & Biederman L (1993) Shy bold continuum in pumpkinseed sunfish (*Lepomis gibbosus*) –An ecological study of a psychological trait. *Journal of Comparative Psychology* 107(3):250-260.

- Wilson DS, Clark AB, Coleman K, & Dearstyne T (1994) Shyness and boldness in humans and other animals. *Trends in Ecology & Evolution* 9(11):442-446.
- Wey, Tina; Blumstein, Daniel T.; Shen, Weiwei; Jordan, Ferenc (February 2008) Social network analysis of animal behaviour: a promising tool for the study of sociality, *Animal Behaviour*, Volume 75, Issue 2, February 2008, Pages 333-344, ISSN 0003-3472, DOI: 10.1016/j.anbehav.2007.06.020.
- Wong, S. V.; Hamouda, A. M. S. (February 2002)Development of genetic algorithm-based fuzzy rules design for metal cutting data selection, *Robotics and Computer-Integrated Manufacturing*, Volume 18, Issue 1, February 2002, Pages 1-12, ISSN 0736-5845, DOI: 10.1016/S0736-5845(01)00019-9.
- Yang, Heng-Li; Wang, Cheng-Shu (August 2008) Two stages of case-based reasoning - Integrating genetic algorithm with data mining mechanism, *Expert Systems with Applications*, Volume 35, Issues 1-2, July-August 2008, Pages 262-272, ISSN 0957-4174, DOI: 10.1016/j.eswa.2007.06.027.
- Yang, Wan-Shiou; Dia, Jia-Ben (April 2008) Discovering cohesive subgroups from social networks for targeted advertising, *Expert Systems with Applications*, Volume 34, Issue 3, April 2008, Pages 2029-2038, ISSN 0957-4174, DOI: 10.1016/j.eswa.2007.02.028.
- Yuan, Yufei; Zhuang, Huijun (November 1996) A genetic algorithm for generating fuzzy classification rules, *Fuzzy Sets and Systems*, Volume 84, Issue 1, 25 November 1996, Pages 1-19, ISSN 0165-0114, DOI: 10.1016/0165-0114(95)00302-9.

Zadeh, L.A. (1975) "The concept of a linguistic variable and its application to approximate reasoning, Parts 1, 2, and 3," Information Sciences, 1975, 8:199-249, 8:301-357, 9:43-80.

