2006

A Comparison of Methods to Construct an Optimal Membership Function in a Fuzzy Database System

Joanne Marie Cunningham
Virginia Commonwealth University

School of Engineering
Virginia Commonwealth University

This is to certify that the thesis prepared by Joanne Marie Cunningham entitled A COMPARISON OF METHODS TO CONSTRUCT AN OPTIMAL MEMBERSHIP FUNCTION IN A FUZZY DATABASE SYSTEM has been approved by her committee as satisfactory completion of the thesis requirement for the degree of Masters of Science in Computer Science

Dr. Lorraine M. Parker, Associate Professor of Computer Science, School of Engineering

Dr. David Primeaux, Associate Professor of Computer Science, School of Engineering

Dr. Amita G. Chin, Associate Professor of Information Systems, School of Business

Dr. David Primeaux, Interim Chairman, Department of Computer Science, School of Engineering

Dr. Robert J. Mattauch, Dean, School of Engineering

Dr. F. Douglas Boudinot, Dean of the School of Graduate Studies

May 2, 2006

A COMPARISON OF METHODS TO CONSTRUCT AN OPTIMAL MEMBERSHIP

FUNCTION IN A FUZZY DATABASE SYSTEM

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science at Virginia Commonwealth University.

by

JOANNE MARIE CUNNINGHAM
Bachelors of Science in Computer Science, Virginia Commonwealth University, 2004
Bachelors of Arts in Religious Studies, Virginia Commonwealth University, 2004

Director: DR. LORRAINE M. PARKER
ASSOCIATE PROFESSOR OF COMPUTER SCIENCE

Virginia Commonwealth University
Richmond, Virginia
May 2006

Acknowledgement

I would like to thank Dr. Lorraine M. Parker for her inspiration and assistance, which was integral to the completion of this project. I would also like to thank her for her interest in furthering the success of women in Computer Science.

I would like to thank all the faculty and staff in the Computer Science Department for sharing their wealth of knowledge, and especially Ms. Deanna Pace for her constant encouragement.

A special thanks goes to my friends inside and outside the department who have provided encouragement and support throughout this research.

Finally a most special thanks goes to my family for their constant love and their support of this research and my entire education.

Table of Contents

List of Tables

List of Figures

Page

Abstract

A COMPARISON OF METHODS TO CONSTRUCT AN OPTIMAL MEMBERSHIP

FUNCTION IN A FUZZY DATABASE SYSTEM

By Joanne Marie Cunningham,
Bachelor of Science in Computer Science
Bachelor of Arts in Religious Studies

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2006

Major Director: Dr. Lorraine M. Parker
Associate Professor of Computer Science

A fuzzy set is one in which membership in a category is not Boolean, rather items have a degree of membership. Fuzzy databases expand on this idea by storing fuzzy data and allowing data to be retrieved based on its degree of membership. Determining the degree of membership that satisfies the largest number of users is difficult. Five different methods of determining the membership function: the Direct Rating Method, the Random Method with step sizes of .02 and .03, the Steplock Method, and the Weighted Average Method, were compared on the basis of convergence and user satisfaction. The results support use of the Direct Rating Method and the Steplock Method in conjunction with each

other, to produce the membership function in the least time and with the highest user satisfaction.

CHAPTER 1 INTRODUCTION

1.1 Fuzzy Data

It is easy to represent hard data, that is data that is specific and not subject to vagueness, e.g. everyone agrees that a rock is not an animal. However, representing data which is not hard in nature; such as whether a bacteria an animal, is more complex. Some people may consider a bacteria to be an animal, thus it becomes necessary to define to what extent a bacteria is an animal. Zadeh introduced fuzzy sets in 1965 in an attempt to classify data that does not fall directly into sets [14].

In classical sets, an element is mapped onto a set with a characteristic function ($f_A(\chi)$) which takes the values $\{0,1\}$. Using this definition an element either belongs to a set (1) or does not belong to a set (0). Consider the example of the rock, a rock is not an animal thus it has no belonging in the category of animal. The characteristic function value for a rock being an animal would be 0. However a rock is something which in non-living, thus it fully belongs to the set of inanimate things, giving the characteristic function for a rock as an inanimate object a value of 1. Figure 1 shows a sample characteristic function for the set animal.

Figure 1. A Characteristic Function for the Set Animal

$$f_{\text{Animal}}(x) = \begin{cases} 1 & \text{if } x \in \text{Animal} \\ 0 & \text{if } x \notin \text{Animal} \end{cases}$$

In fuzzy sets it is necessary to define a degree to which something belongs in the set. The example of the bacteria illustrates this since bacteria has some degree of belonging in the set animal. The characteristic function used for classical sets falls short in this area, thus it is necessary to define a different function to model belonging to a fuzzy set.

1.2 The Membership Function

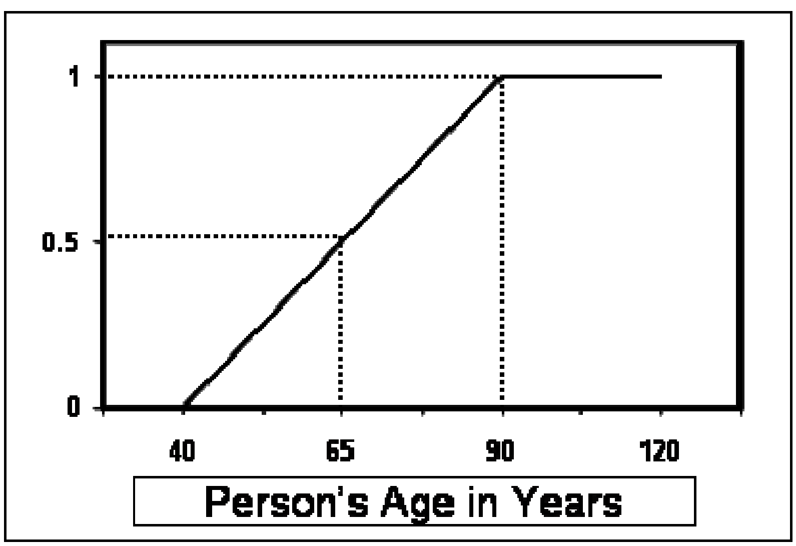
In fuzzy sets, a membership function $\mu F(x)$ is used to map an item onto the interval [0,1]. The value of the membership function, or weight, is the extent to which an element belongs to a set [14]. The membership function below defines the degree to which a person belongs to the set old:

Figure 2. A Membership Function for Old

$$\mu_{\text{Old}}(x) = \begin{cases} 0 & \text{if } 0 < \text{Age}(x) \leq 40 \\ 0.01 \text{ - } 0.99 & \text{if } 41 \leq \text{Age}(x) < 90 \\ 1 & \text{if } 90 \leq \text{Age}(x) \end{cases}$$

A person who is 41 would have a lower degree of membership in the set of old people than a person who is 80. The sample membership function in Figure 2 above maps a person, whose age is less than or equal to 40, to 0, meaning that person has no membership in the category old. Someone in the age range of 41 to 90 would get some degree of membership in the category of old with this value increasing as the age increases. For a person age 90 or above the membership function is 1, meaning the person has full membership in the category old. Figure 3 graphically displays this membership function as a gradual transition between no membership and full membership in the category old.

Figure 3. Membership Function for Old



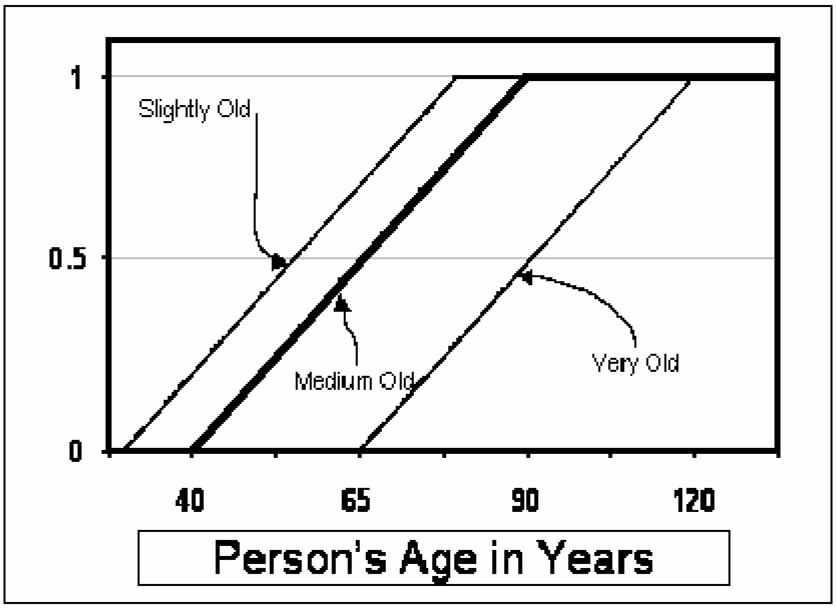
While it may appear that this degree of membership is the probability that a person is old, this is not the case. Membership in a fuzzy set is not a statistical value.

1.3 Linguistic Modifiers

Linguistic modifiers, also known as hedges, are adjectives such as slightly, medium, very, more, dark, light, and extremely, which are used to further define the membership function. Other words such as technically, almost, essentially and practically can also be used as hedges. The effect of this second set of words on the membership function is more complicated and is dependant on the context of the membership function which they are modifying. For this reason, modifiers are separated into two categories the first list of words are Type I modifiers and the second list are Type II modifiers as described by Zadeh [15].

Thus far atomic membership functions have been discussed; however there exist composite membership functions which result from the concatenation of a linguistic hedge with a fuzzy set. Thus slightly old is a valid membership function as are medium old and very old. When membership functions are constructed using modifiers the values of the membership function are shifted as shown in Figure 4.

Figure 4. Membership Functions for Old with Linguistic Modifiers



In this case the modifier slightly shifts the membership function down such that a person age 65 has full membership in the category old. The membership function for very old is shifted to the right such that a person must be over age 100 before they have full membership in the category.

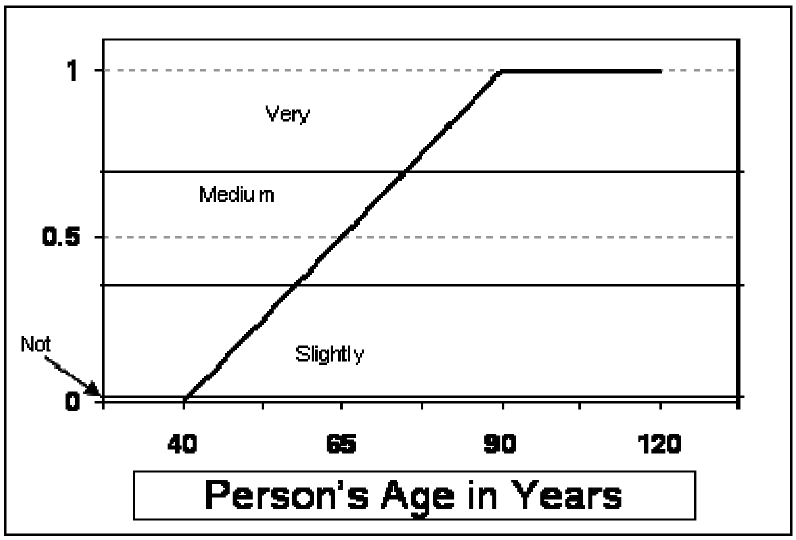
Because linguistic modifiers have this effect on the membership function they can be used to describe the degree of 'oldness'. With this use of modifiers, the membership function is modified such that the weight of the membership function is mapped to a term, which exemplifies the degree of membership. Thus instead of having a person be old or not, with the use of modifiers a person can be categorized as slightly old, old or very old. If the weight (the degree of membership) is lower, the person is put into the slightly old category, if the weight is higher the person is put into the old or very old category. An example of weights with linguistic modifiers is shown in Table 1.

Table 1. Linguistic Modifiers and Corresponding Ranges

Linguistic Modifier	**Weight Range**
Not	0.0 - .02
Slightly	0.0 – .35
Medium	0.36 – 0.69
Very	0.7 – 1.0

Notice that Not is included as a modifier in this table. The inclusion of not allows for the possibility that something does not belong to the set or that it has no membership in a set. The necessity of this modifier was experimentally determined in [4]. Using linguistic modifiers with the set old, divides the fuzzy set old into several fuzzy subsets: not old, slightly old, medium old, and very old. This is graphically represented in Figure 5.

Figure 5. Membership Function for Old Split by Linguistic Modifiers



The example shows a scheme where the membership function is split into exclusive sets, however this does not have to be the case. Fuzzy sets have the ability to overlap. Consider a scheme where there is a fuzzy set for old and one for young. A person could have membership in both sets, e.g. be slightly old and slightly young, depending on ranges set for the modifiers. It can be assumed that as membership in one set increases membership in the other set would decrease, e.g. someone classified as very old would most likely fall into a lower grade of membership in the category young as slightly or not young.

CHAPTER 2 DETERMINING THE VALUE OF THE MEMBERSHIP FUNCTION

2.1 Fuzzification and Defuzzification

Fuzzification is an operation which can be performed on a non-fuzzy or fuzzy set to make the set more fuzzy. The operator ~ is a fuzzifier and represents that a set has been made more fuzzy. Take for example the non-fuzzy value 5. Thus the fuzzy value $\underset{\sim}{5}$ is the set of numbers which are approximately equal to 5. Fuzzification can also be done on operators where the $\underset{\sim}{=}$ operator is approximately equal to and $\underset{\sim}{<}$ is approximately less than. When eliciting the membership function from a community of users, fuzzification occurs as the users are polled for their input regarding the membership function [15].

The process by which a crisp value is chosen to be indicative of the fuzzy membership function is called Defuzzification [9]. Methods of defuzzification include: Maxima Methods, methods which select the membership function with the maximum; Distribution Methods, methods which compute a probability distribution then select a value based on probability (this includes center of gravity calculations); and Area Methods, where the area under the membership function is used to determine the value of the membership function (this includes a center of area calculation). Other miscellaneous methods can be used to determine the value of the membership function, such as clustering. The performance of a method used for defuzzification is dependant on the data and desired result of the system [9]. After users from a community have been polled, defuzzification is used to construct an optimal membership function which is returned as a crisp value.

2.2 Constructing the Membership Function

Determining the value of the membership function is not an exact science. Consider the previous example of “old”. A scheme could be created that would always map a person under 40 to a low degree of membership in the category old. However, a person who is 8 might say that a 40-year-old is very old. Thus there is an element of fuzziness that comes from the community perceiving the data.

There are several ways of determining the membership function. The method used depends on the desired behavior of the system and the designer’s view of a membership function. The validity of the value of the membership function is highly dependant upon the user community of the system. Thus, it is important that the membership function be consistent with the perceptions of the users of the system. Tashiro [13] proposes the idea of defining two membership functions in a fuzzy database. The first is a universal membership function for all users, while the second is a membership function defined specifically for each individual user. These membership functions are used in combination to cater the results of a query to a given user [13].

In the VCU fuzzy database system it is desirable that a single membership function return the same value for each query regardless of the user. This requirement enables the system to be trained by users that are representative of the final user community, but are not necessarily the final users, eliminating the need for each user to train the system individually. Thus the membership function must be representative of the views of the majority of the users.

There are various ways of defining this membership function. The database designer can separate the elements, fuzzy items which are stored in the database, into fuzzy sets and associate weights with the elements. This gives the developer's perspective of how the user community would set the weights. This method of defining the membership function is undesirable because it does not take into account the opinions of the community of users.

Another way to determine the function is to elicit information from the community of users during a training phase. During this phase individual members of the user community are asked for their opinion regarding some fuzzy aspect of the set, for example, how old is Joe? Fuzzification occurs during this training phase as several opinions about the value of the membership function are obtained. The opinions are then used to construct the membership function stored in the database through defuzzification. The training is considered complete when some form of convergence criteria is reached.

2.3 Methods of Eliciting the Membership Function

There are several ways that information can be presented to the user in order to solicit data which is used in the construction of the membership function. Bilgic and Turksen [1] discuss six different methods of questioning the user in order to gain information and build a membership function. The following is a summary of their methods:

2.3.1 Polling

The polling method stems from the idea that fuzziness is a result of disagreements between individuals. In polling, multiple people are asked a question of the general form, "Do you agree that element x is classification y?" For example, when classifying a person (Tom) as old or young, the question would be "Do you agree that Tom is old?" Answers to this question are used to create a membership value for Tom that best agrees with the majority of users. This method for determining the membership function works well with the likelihood interpretation of the membership function, which says that $\mu F(x)$ (the value of the membership function) equals the percentage of people who said that Tom is old.

2.3.2 Direct Rating

The direct rating method is inspired by the idea that fuzziness results from an individual's inability to definitively assign an element to a category. This rating method requires the same individual to answer the same question multiple times. The training session is carefully designed so that the individual is not likely to remember their previous responses to the questions. The questions are asked in the format how y is x, or "How old is Tom?"

The user then selects from a set of possible values, in this case slightly old, old or very old. This method can be improved by asking several users the repeated questions many times. The membership function is then constructed based on the frequency of particular responses.

2.3.3 Reverse Rating

In the reverse rating method, an individual is presented a degree of membership and asked to pick those elements that fit the membership value. The question asked would be, "Out of these people, which are old?" This same question can be presented to the same user multiple times, as in the direct rating method, or presented to multiple users, or both. The votes are recorded and the value of the membership function is constructed based on all the votes [3].

2.3.4 Interval Estimation

The interval estimation method of eliciting the membership function is based on the idea that the membership function represents the percentage of a population that feels an element x is in the category y. For example a membership value of .75 for Tom being old represents that 75% of the population says that Tom is old. A sample question would be presented in the format "Give the interval on which Tom's age falls." In this case the answers would be old or young. Linguistic modifiers could be included to make the set more descriptive. The membership function is then constructed based on how many people put Tom into which category. This method is especially useful when the attribute in question is measured linearly such as age, height, or temperature. Chameau and Santamarina [3] report that this method has advantages over methods such as direct rating and polling where the user responds with a crisp yes or no answer. They also report that this method produces membership functions that have narrower spread (are more precise) than methods such as direct rating and polling.

2.3.5 Membership Exemplification

The membership exemplification method of determining the membership function is most like the example of the database designer assigning values without polling a group. In this method a person is asked question of the general form, “To what degree does element x belong to category y?” The specific form of this question for the case of Tom’s age is, “To what degree is Tom old?” If the question is only asked to one person, as in the research of Hersh & Carmazza [7] the membership function is simply the value given by the subject. Because this question is asked to a sole user, they report that this method produces a membership function that varies from that obtained by polling or direct rating. This further demonstrates the need for community involvement in determining the membership function. This method of elicitation can be used in a community setting where some function is used to combine the results and construct the membership function.

2.3.6 Pairwise Comparison

In pairwise comparison multiple users are given two elements, x and y, and asked which element is more z and by how much. If we have Bill and Tom the question becomes “Who is older, Bill or Tom, and by how much?" The results of these questions are recorded, combined and the membership function is created.

2.4 Summary

Fuzzification and defuzzification are important steps in constructing a membership function, which encompasses gaining user input and extracting a value of the membership function from that input. Sections 2.3.1 – 2.3.6 discuss multiple methods of eliciting information to construct membership functions. Modified forms of these methods were used in this experiment to gain user feedback and construct a membership function from the feedback.

CHAPTER 3 VCU FUZZY DATABASE PROTOTYPE

3.1 Current Fuzzy Database System

The current database designed by the Database Research Group at VCU contains information about eye color [5]. Eye colors are categorized as Blue, Green or Brown. Within these categories the color is further categorized with the linguistic modifiers Slightly, Medium, and Very. Thus two tables are used to return information stored in the database. One contains the membership function for the eye color, Blue, Green or Brown (Table 2), and one contains the ranges of the membership function for which the given linguistic term applies (Table 3). This table also includes a value for the midpoint of the range. Note that the values chosen are not true midpoints, rather values near the center for that modifier range. For example, the midpoint for the modifier "Not" is set to 0 while the actual midpoint of the range is .01. The value 0 was chosen because it gives a better representation of the concept of "Not."

Table 2. Membership Values (Weights) for Each Attribute of Image 1

IMAGE_ID	EYE_COLOR	WEIGHT (μ)
1	GREEN	0.8
1	BLUE	0.3
1	BROWN	0.0

Table 3. Modifier Ranges and Midpoints

Modifier	Range_From	Range_To	Midpoint
"Not"	0.00	0.02	0.00
"Slightly"	0.03	0.35	0.20
"Medium"	0.36	0.69	0.52
"Very"	0.7	1.0	0.85

The information in Table 2 represents an image with very green eyes and medium blue eyes. Previously, the modifier "Not" was not included in the system [12]. Thus the range of slightly was from 0.00 to 0.35. This created the problem that if the system was queried for people with slightly brown eyes this image was returned as having slightly brown eyes, although it is clear that the person in the image does not have brown eyes. Thus the previous scheme lacks a way to specify that an image does not have brown eyes. “Not” was added as a modifier in the current version of the prototype to solve this problem.

In the current database, information is retrieved using queries on both the eye color and the linguistic modifier. The queries are constructed in the format:

```
SELECT (attribute list)
FROM (relation list)
WHERE (fuzzy conditions)
```

SQLf, a modified query language, which introduces fuzziness into queryies is used on top of SQL Server to provide the extra querying capabilities needed for fuzzy query processing. Thus a query such as:

```
SELECT IMAGE_ID
FROM Color
WHERE  EYE_COLOR = SLIGHTLY BLUE
```

can be expressed and processed. Additionally a natural language interface has been implemented which can parse queries in the form “Give me all the people with slightly blue eyes” [2].

3.2 Previous Research Contributions

Research has been conducted as to the best way to initialize the membership function along with the best way of eliciting information from users to define the membership function. The work of Lee [8] compares several different methods of initializing the membership function weights to determine which will most quickly lead to convergence in the training phase of a fuzzy database system. In this study convergence is defined as the point where additional feedback from the user community have no effect on the value of the membership function. In this case the user was the researcher and thus once the images were categorized correctly according to the researcher, training was complete. These methods were evaluated by the speed of convergence, and Lee concludes that the best way to initialize the weights is in the midpoint of the modifier range. However she states a concern that the stability of convergence within a community should be addressed in future research.

Research by Sanghi tested various methods of eliciting the value of the membership function to determine which produced a membership function with a higher degree of user satisfaction [12]. The Random Method was compared to the Direct Rating Method. For the training session, in which the membership function was determined, the Random Method membership weights are initialized to a random value between 0 and 1. Users are then shown pictures that meet criteria such as EYE_COLOR = SLIGHTLY BLUE. The users provided feedback on the color and if a user supported a weaker modifier the weight was decreased by 0.01. Likewise, if a stronger modifier was supported the weight was increased by 0.01.

In the Direct Rating Method the users were asked to rate the eye color as green, blue, or brown on a sliding scale. A frequency distribution function was created and the membership weight was set to whatever category within each color (e.g. slightly green, very green, or medium green) had the highest number of votes. As stated previously, this scheme does not allow for a picture to be categorized as having eyes that are not a color (e.g not green).

After membership functions were obtained for both methods, a testing session was conducted in which users were asked how satisfied they were with the result. The goal was to obtain a membership function with the highest degree of user satisfaction. Sanghi found that the Direct Rating Method produced a higher degree of user satisfaction than the Random Method.

In addition a system is under development by Mastros [10] and McDermott [11] in which spatial information about images is stored with regards to nose length. The fuzzy attribute "length of nose," is categorized as short, medium, or long. The length of the nose is known to the researcher and the initial membership function value is based on this length. The value of the membership function is then changed based on user input. In this research it would be desirable if the actual length of the nose corresponded in some way to the value of the membership function after training. This research is of particular interest because of the development of an alternative Steplock Method of changing the weights associated with the length of the nose [10].

3.3 Methods of Modifying the Membership Function

Four different methods of modifying the membership function have been developed for use in determining the membership function in the VCU fuzzy database system. These are the Direct Rating Method, the Random Method, the Steplock Method, and the Weighted Average Method.

3.3.1 Direct Rating Method

The Direct Rating Method was used in tests by Sanghi [12]. The goal of this method is to place an image in the category in which the most users place it. It makes use of the polling method of eliciting the membership function discussed in section 2.3.1. For this method users are asked whether an image has eyes that are slightly, medium, or very and a color. For example images would be displayed and a user would be asked: How Blue are these eyes? They are given the choices: Slightly Blue, Medium Blue, Very Blue or Not Blue. The number of votes for each modifier are counted and the weight is set to the midpoint of the modifier range which had the most votes. This is repeated for each color and each image.

This method does not need to be initialized; however, other problems are introduced. With this method it is possible to have a situation where two categories have nearly the same number of votes and there are additional votes for a different category. Thus the membership weight chosen may not be suitable for the majority of users of the system.

3.3.2 Random Method

Lee [8], Sanghi [12] and Mastros [10] all use a version of the random method in their research. The goal of this method is to place all images in an initial category so that if necessary, queries can be run against the database before a training phase is conducted. With this method users gradually move images to an appropriate category. In initial versions of this method membership values are randomly initialized (yielding the name Random Method). Lee [8] discusses other methods of initializing the membership values including the Random Proportional Method, New Random proportional method, and Midpoint methods.

In the Random Proportional method the membership value for one color is set to a randomly generated value M. The membership value for the other colors was set to (1 - M) / 2. This method is inspired by the idea that it is unlikely that all 3 colors have the same membership values rather the values are proportional. The goal of this method is to speed convergence. In the Random Proportional Method, the same color is always set to a value first; it was thought that this might create some bias towards the first color set. Thus the New Random Proportional Method was developed in which the first color is rotated between the three colors. In the Midpoint Method, the weights are initialized to the midpoint of the possible weight range, i.e. 0.5. This method subscribes to the idea that at the midpoint it will take equal time to move to high or low extreme values. In Lee's comparison of these methods, she found that the Midpoint Method best facilitated convergence in a single user system [8].

After initialization, images are displayed for each color / modifier pair. For example, all the images with slightly blue eyes will be displayed. The users are then asked how well the images meet the criteria. Answer Choices are: Meets Criteria, Less Blue, More Blue, or Not Blue. This questioning technique was adapted from the polling method of eliciting the membership function described in section 2.3.1. If the vote is Meets Criteria, the weight is moved y steps either up or down towards the midpoint. If the vote is for a higher category the weight would be increased by y. If the vote is for a lower category, the weight would be decreased by y. If it is for the current category it is moved towards the midpoint of that category by y.

This method of modifying the weights is faulty because it linearly changes the weight of the images. This method is not robust against data bursts. For example, in the current implementation, if 300 users say that eyes are Slightly Blue another 30 (or so, depending on the step size used to change the weights) users voting for Very Blue could change the weight of the Blue attribute to be Very Blue, disregarding the fact that the majority of the users believe that the eyes are Slightly Blue. Additionally when using this method the issue of how to appropriately initialize the membership function is raised.

3.3.3 Steplock Method

The Steplock Method was developed and tested by Mastros [10]. The goal of this method is to prevent the input of a few users from undoing the opinion of a larger group of previous users. The initial weights for each color are initialized the same as they were for the Random Method. Questions are of the same format as they are for the Random Method. However, the effect of votes is different. If a vote is for the same range as the current weight, 1 is added to the step size out of that weight. If a vote is outside of the range and the current step size is greater than 0 then the step size is decreased by 1. When the step size is 0 and a vote is outside of the range, the weight is increased or decreased by .03 in the direction the vote indicates.

By adding steps, this method makes it more difficult to change a weight that has been voted in the same category by many users. This method is faulty because it does not take into account all votes placed over time. Votes are essentially thrown away after steps are removed, so there is no record of several users having voted the same way for an image. The same issue of initialization is valid for this method as it is for the Random Method.

3.3.4 Weighted Average Method

The weighted average method was developed to address the issues of the current methods of modifying the weights [4]. The goal of the weighted average method is to include all of the previous votes into the weight calculation as opposed to throwing away votes after they have been used to calculate the weight. Questions are asked in a similar format to the Direct Rating Method. In this method the number of votes for each weight is counted. This number is multiplied by the midpoint for that particular weight. The sum of these calculations is then divided by the total number of votes to get the current value of the membership function. A sample calculation for the Blue weight would be:

$$\mu F(x) = \frac{\text{(Votes Slightly)(.2) + (Votes Medium)(.52) + (Votes Very)(.85) + (Votes Not)(.0)}}{\text{Total Votes}}$$

3.4 Preliminary Study

A preliminary study was conducted to test the feasibility of the current methods along with the proposed weighted average method. 27 images were showed to 29 users for evaluation. The user feedback was processed and membership functions were constructed for each image using each of the four methods for determining the membership function. In this study nearly all the methods placed the images in the same category. The Direct Rating and Weighted Average methods had nearly the same number of users before reaching their final weight. The Steplock and Random Methods took more user feedback to achieve their final weight.

From this study it was concluded that the Weighted Average method is a viable option for determining the membership function. However, were needed to determine the effects of a larger user base on the convergence and robustness of the membership function. Finally, a testing phase should be conducted to determine which method produces a membership function which pleases the most users [4].

3.5 Summary

The previous work by the Database Research Group set the stage for this current experiment. Previous research left several unanswered questions regarding how to best construct the value of the membership function, as no work was done comparing all methods of modifying the membership function. It is the goal of this research to compare these different methods and develop some conclusions as to the best way to construct the membership function.

CHAPTER 4 CURRENT PROJECT

4.1 Goals

The goal of this project is to determine the best method of collecting and processing user opinions to construct an optimal membership function. In future implementations of the Fuzzy Database System this information will be used to specify which method of determining the membership function should be used.

4.2 Methodology

An application similar to those used in prior experiments was developed, however this application was updated to work with multiple methods of modifying the membership function and written in C#. Minor changes were made to the user interface for this application. Six images per page were shown to users and they were asked various questions about the images, depending on which method of modifying the membership function was being used. A total of 27 images were shown to each user for each method of evaluating the membership function.

Both a training phase and a testing phase were conducted. During the training phase users were asked their opinions of the eye color of images. This feedback was evaluated using multiple algorithms to determine the membership function, and membership functions for each image were constructed.

After the training phase a testing phase was conducted. In the testing phase images were presented to users in the category which they had been placed during the training phase. The category was the modifer range in which the weight of the membership value fell. The ranges for modifiers were as shown in Table 1, with the exception of the Weighted Average "Slightly", and "Not" categories. For "Not" the upper bound of the range was changed to .10, in accordance with the finding that increasing this value increases accuracy with this method [4]. Additionally, because of this change the lower bound for "Slightly" was changed to .11. The users were asked whether they felt the image had been placed in the appropriate category.

4.3 Experimentation

Five different methods of determining the membership function were evaluated in this experiment: the Direct Rating Method as described in section 3.3.1, two versions of the Random Method as described in section 3.3.2; the Steplock Method, as described in section 3.3.3; and the Weighted Average Method as described in section 3.3.4. The two versions of the Random Method were Random(.02) and Random(.03) where the step size (the value that is added and subtracted from the membership function based on the user response) was .02 and .03 respectively. These two step sizes were chosen because a larger step size is expected to move to a final answer sooner, however a smaller step size may create a membership function which is more robust.

In the training phase, for each method of modifying the membership function, images were displayed to the users and the users were asked a question about the image. For the Direct Rating and Weighted Average Methods, all images were displayed with each color (green, blue, brown) and the user was asked, “How <color> are these eyes?” Answer choices were: “Slightly <color> Eyes,” “Medium <color> Eyes,” “Very <color> Eyes,” and “Not <color> Eyes.” For the Random(.02), Random(.03) and Steplock Methods, images currently in a category were shown to a user. There were 9 categories, one for each color (green, blue, brown), modifier (slightly, medium, very) pair. Images in the not category were returned with the images in the slightly category as had been done in previous research [8]. The user was told the category and asked their opinion about the eye color. For example: “Here are people with <modifier>, <color> eyes.” Answer choices were: “Meets Criteria”, “More <color>”, “Less <color>”, “Not <color> Eyes.” This was conducted with a minimum of 65 and maximum of 117 users per image.

After the training phase was completed, a testing phase was conducted. During the training phase the images were placed into a category. This category was potentially different based on the method of modifying the membership function used. In the testing phase, the images, which had been placed in each category (color, modifier pair, including not), were displayed to the user. The user was then asked whether or not they were satisfied with the image in that category. A sample question was, "People with <modifier> <color> Eyes," and answer options were "Satisfied," "Not Satisfied." This experiment was conducted with either 50 or 51 users per image. The overall satisfaction rating was the sum of the positive votes for each color, divided by the total votes the image received. For example if Image 10 was in the Slightly Blue category, Medium Green category, and Not Brown category, the calculation was:

$$\textit{Satisfaction} = \frac{SV(\textit{Slightly Blue}) + SV(\textit{Medium Green}) + SV(\textit{Not Brown})}{\textit{Total Votes}}$$

Where SV(x) is the number of satisfied votes for a category x.

4.4 Evaluation

The criteria for evaluating the membership functions were the length of time to get to the final membership function, robustness of the membership function and user satisfaction with the membership function. The length of time to obtain the final membership function was defined as the number of votes needed to move the image into the last category it was moved to, which was assumed to be the best possible category. The robustness of the algorithm used to calculate the membership function was also evaluated using this metric. The weight for number of votes was graphed to visualize robustness. User satisfaction with the final membership function was evaluated in the testing phase. The user satisfaction for each image was compared for different methods of modifying the membership function.

CHAPTER 5 RESULTS

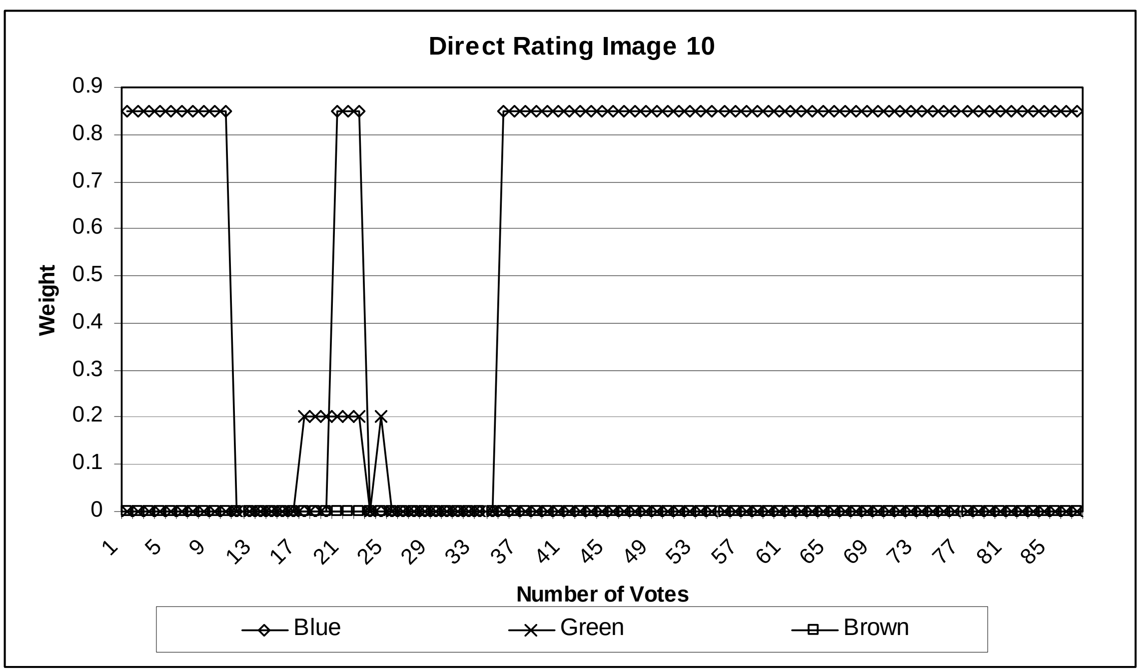
5.1 Direct Rating Method

In the Direct Rating method 87 users were questioned for each image. The number of votes needed to obtain the final value of the membership function for each color is summarized in Appendix A Table 5. Each color is calculated separately then an average number of votes for each color was taken to determine the average number of votes needed for this method. The number of votes needed for each attribute to reach a modifier which remained the same through the rest of the voting was defined as the number of votes needed to reach the final value or modifier. This number was obtained by counting the number of votes obtained before the modifier no longer changed. The votes needed for each attribute to reach its final value ranged from 1 to 86. The average number of votes needed for the blue attribute to reach its final value was the highest at 16.70 votes followed by green with 9.88 votes and brown needing 1.63 votes.

Assuming that all colors must reach their final value before the image has reached convergence, it was necessary to create an additional overall column for each image. This column contains the number of votes needed for the color that took the maximum number of votes for each image. The average of this column was used for comparison purposes between the methods. The average overall number of votes needed for the Direct Rating Method was 23.19.

Because the value of the membership function is set to the midpoint of the current modifier, the value of the membership function remains constant over time until the number of votes for another category exceeds the votes for the current category and the value of the membership function is changed. This trend is shown in Figure 6 where the weight for each attribute is graphed over the number of votes obtained for sample image 10.

Figure 6. The Direct Rating Method for Image 10



The percentage of users satisfied with the classification of each image is summarized in Appendix B Table 10. Fifty or 51 votes were gathered per image in the testing portion of this experiment. The percent of users satisfied with the classification of each image ranged from 60.93% to 91.39% with an average of 79.29% for the Direct Rating Method.

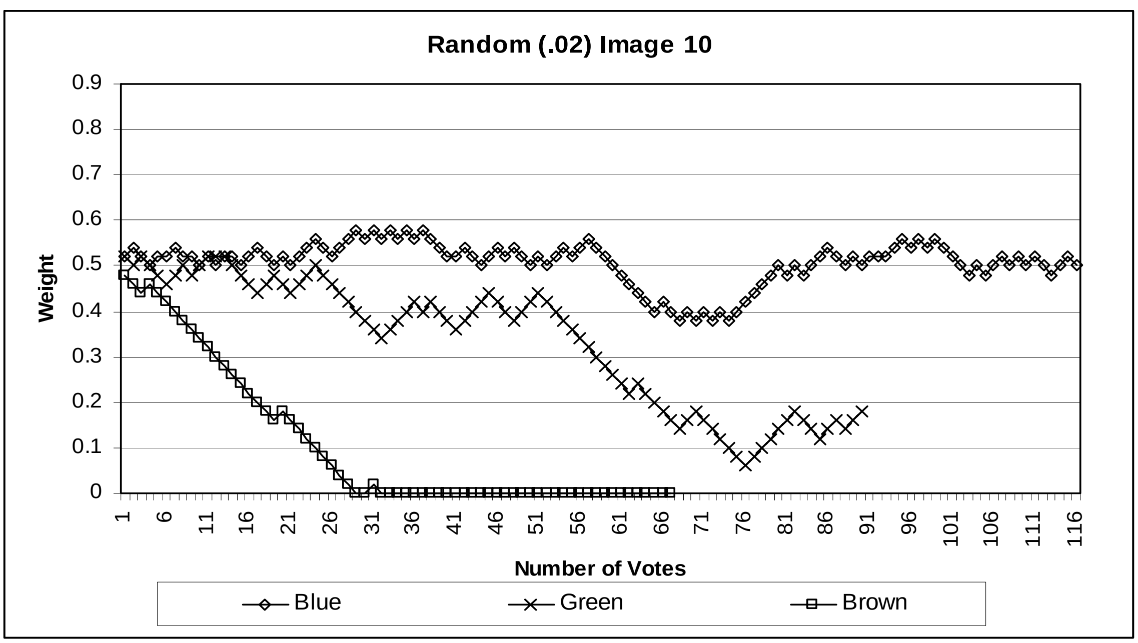
5.2 Random Method (Step Size .02)

For the Random Method with a step size of .02 between 65 and 117 users were polled for each image during the training phase. The initial query for this range did not operate properly and thus not all images were returned for users to vote on. Additional trials were done with this method to obtain a suitable number of users for all images. The number of votes that were required for each image to reach its final weight for each color is summarized in Appendix A Table 6.

The votes needed ranged from 0 to 116 for an image to reach its final modifier in a color category. Zero votes are needed if the weight is initialized to a value that is within the range of the final modifier. For example, the final modifier is medium and the value never is moved from the medium category then the number of votes needed would be 0. The average number of votes needed for the blue attribute was the lowest, 22.96 followed by the brown attribute at 39.63 then the green attribute at 51.89 votes. In most cases where the number of votes was high the image was moved from the not category to the slightly category for a vote then moved back to the not category. The average overall number of votes needed for each image was 68.11.

Figure 7 shows the weight of each color graphed over the number of votes obtained for image 10. In this chart, different numbers of votes were obtained for each attribute, resulting in different lengths of the lines for each attribute. As shown, the weight for the Random Method moves linearly by a factor of .02 as users vote. The line for brown shows an initial drop for the image to be moved to the not brown category. This initial movement adds to the number of votes needed to achieve the final value of the membership function. The value for green moves more slowly to the slightly category, and the line for blue maintains consistently in the medium category.

Figure 7. The Random (.02) Method for Image 10



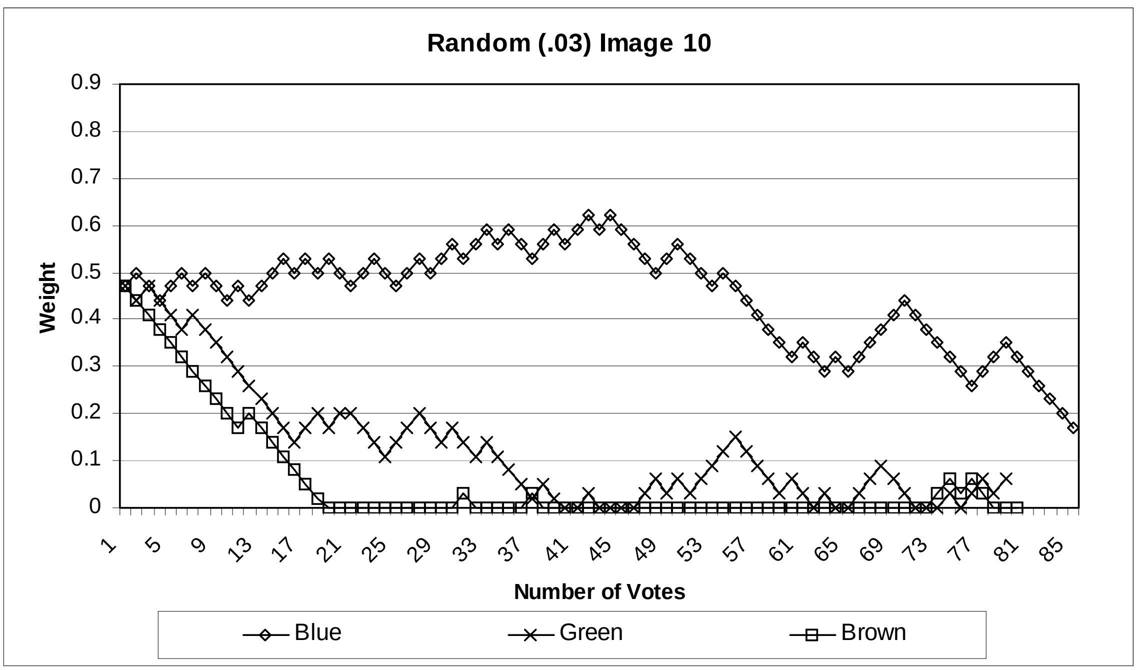
The percentage of users satisfied with the classification of each image is summarized in Appendix B Table 10. Fifty or 51 votes were gathered per image in the testing portion of this experiment. For the Random (.02) Method, the percent of users satisfied with the classification of each image ranged from 50.00% to 91.39% with an average of 78.97%. The low satisfaction value was for image 33 which had been placed in the Not Green category by most users; however, the last two users voted and moved it to the Slightly Green category.

5.3 Random Method (Step Size .03)

For the second Random Method implemented with a step size of .03 between 80 and 85 users provided feedback for each image. The number of votes needed for an image to reach a final modifier is summarized in Appendix A Table 7. This value ranged from 0 to 85, with 0 occurring in the same situation as described in section 5.3. The average number of votes needed for blue was 44.26, followed by brown at 59.29, then green at 62.67. In this case, like the Random (.02) method, higher numbers of votes were observed by images in the not category, as the images were moved from the not category to the slightly category then back. More instances of this occurred for this method than the Random (.02) method because the threshold for Not was set at .03 and thus a single vote for Slightly could move the image out of the not category. The average overall number of votes needed per image was 77.74. The value placed in the overall category nearly always came from a color attribute with the final modifier Not.

Over the number of votes the Random (.03) method has a similar curve as the Random (.02) method, however the step size is larger so the slopes are steeper. While the image takes fewer votes to reach an appropriate value, this method does not maintain a value well, especially in the case of the modifier Not. Figure 8 is a graph of this method for sample image 10 displaying these trends.

Figure 8. The Random (.03) Method for Image 10



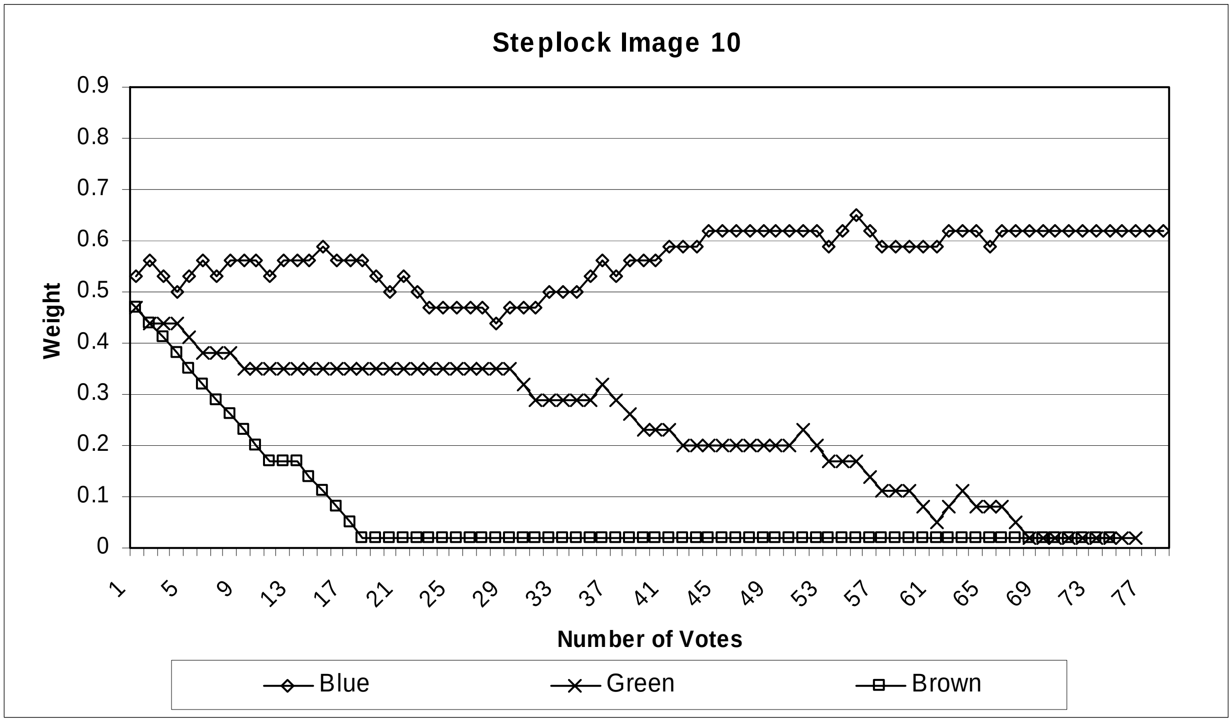
The percentage of users satisfied with the classification of each image is summarized in Appendix B Table 10. Fifty or 51 votes were gathered per image in the testing portion of this experiment. The percent of users satisfied with the classification of each image ranged from 56.67% to 91.39% with an average of 79.54% for the Random (.03) Method. Similarly to the Random (.02) method the low value was for image 33 which had been placed in the not green category by several users, however the last two votes moved it to a slightly category.

5.4 Steplock Method

For the Steplock Method, between 72 and 79 votes were obtained per image. The number of votes needed for an image to reach a final modifier is summarized in Appendix A Table 8. This value ranged from 0 to 68 within each individual color group. Zero occurred when the images remained in the initial category the as described in section 5.3. The average number of votes needed for the image to reach its final membership weight was 18.81 for the blue attribute, followed by the brown category with 21.52 and the green attribute with 23 votes. The average number of votes needed overall, that is for each color to be placed in its final group, was 33.81.

The main goal of the Steplock Method is to maintain a value over a period of time with less fluctuation if votes are for the same category. This method was designed to be robust against data bursts and disagreeing users. This effect is demonstrated in Figure 9, which shows the change in weights as users vote for image 10.

Figure 9. The Steplock Method for Image 10



The horizontal sections of the lines, for example the Green line between votes 13 and 29, are areas where users voted for the current category and the image resisted change.

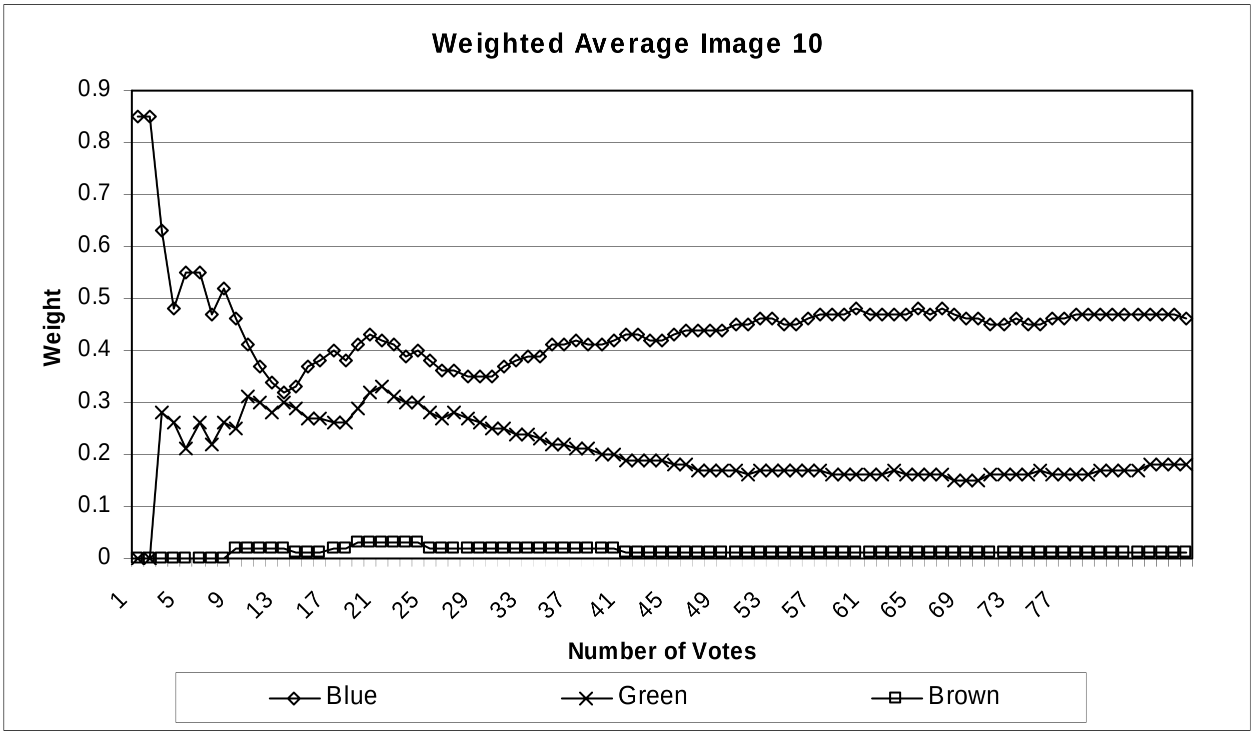
The percentage of users satisfied with the classification of each image is summarized in Appendix B Table 10. Fifty or 51 votes were gathered per image in the testing portion of this experiment. The satisfaction ranged from 63.33% to 91.39% with an average satisfaction rating of 79.66%. This method maintained the Not modifier for image 33, unlike the Random methods. This demonstrates an advantage of the Steplock Method, that the method is robust against changes made by only a few users.

5.5 Weighted Average Method

A total of 87 votes per image were obtained for the Weighted Average Method. The number of votes needed for an image to reach a final modifier is summarized in Appendix A Table 9. The number of votes needed to obtain a final weight for each color with this method ranged from 1 to 85. Green took the fewest votes with an average of 7.44 votes per image followed by Blue with 13.81 votes for image the finally by Brown at 15.40 votes per image. The average overall number of votes needed was 29.37.

Figure 10 shows the graph of the Weighted Average Method over the number of votes obtained. The weight initially jumps around, but then becomes steadier as the number of votes increases and maintains a value over a period of time. The more gradual change in behavior occurs as more votes are collected, because as more votes are included in the average, each vote has less influence on the weight.

Figure 10. The Weighted Average Method for Image 10



The percentage of users satisfied with the classification of each image is summarized in Appendix B Table 10. Fifty or 51 votes were gathered per image in the testing portion of this experiment. The percentage of users satisfied with the category an image was placed in with this method ranged from 48.00% to 91.39%. The average satisfaction was 77.24%. This is slightly less than the other categories and it is important to note that this method often placed an image in the Slightly category when other methods placed the same image in the Not category.

CHAPTER 6 CONCLUSION

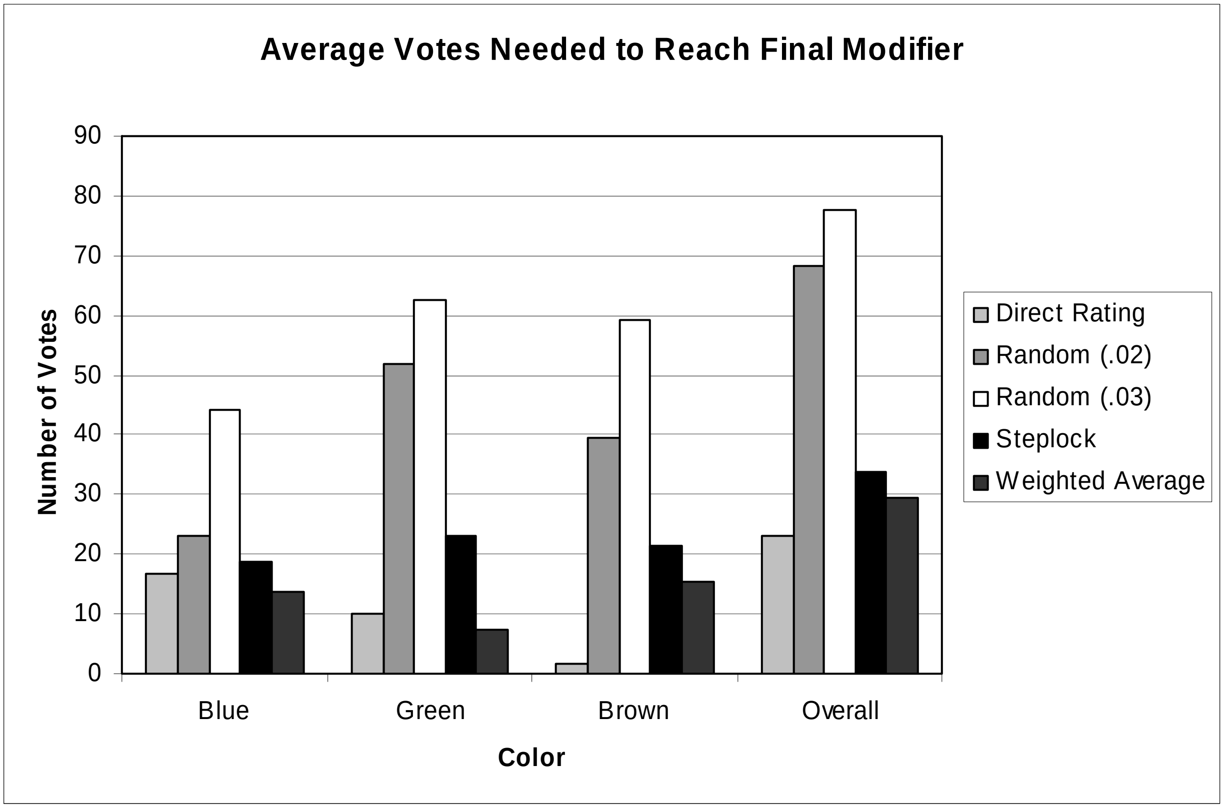
6.1 Comparison of Methods

The results of the different methods used to construct the membership function as described in sections 5.1 through 5.5 were compared on the basis of number of votes needed to reach the final membership function, and the number of users satisfied with the final membership function. The number of votes needed to construct the final membership function for each method, for each color and overall is summarized in Figure 11. Table 4 shows the average number of votes needed to reach the final modifier.

Table 4. Average Number of Votes Needed to Reach the Final Modifier

Method	Blue	Green	Brown	Overall
Direct Rating	16.7	9.89	1.63	23.19
Random (.02)	22.96	51.89	39.63	68.11
Random (.03)	44.26	62.67	59.3	77.74
Steplock	18.81	23	21.52	33.81
Weighted Average	13.81	7.44	15.41	29.37

Figure 11. Average Number Votes Needed to Reach Final Modifier



As shown in Figure 11, the Direct Rating Method took the least number of votes to reach the final modifier for all colors except blue. The final modifier was reached first in the Blue category by the Weighted Average Method. The Random Method with step size of .03 took the most votes to reach a final value, followed by the Random Method with step size of .02 and then the Steplock Method. This is consistant with the findings of Lee [8], that a smaller step size is more appropriate for facilitating convergence in a community, while a larger step size is more appropriate for convergence with one user.

In the overall category, the average number of votes needed for all three of the color attributes to reach their final weight the Direct Rating Method needed the fewest votes. The Weighted Average method took the next fewest, followed by the Steplock method. Both implementations of the Random Method took double the number of votes needed by the Steplock method. With the Random .03 method needing the most votes.

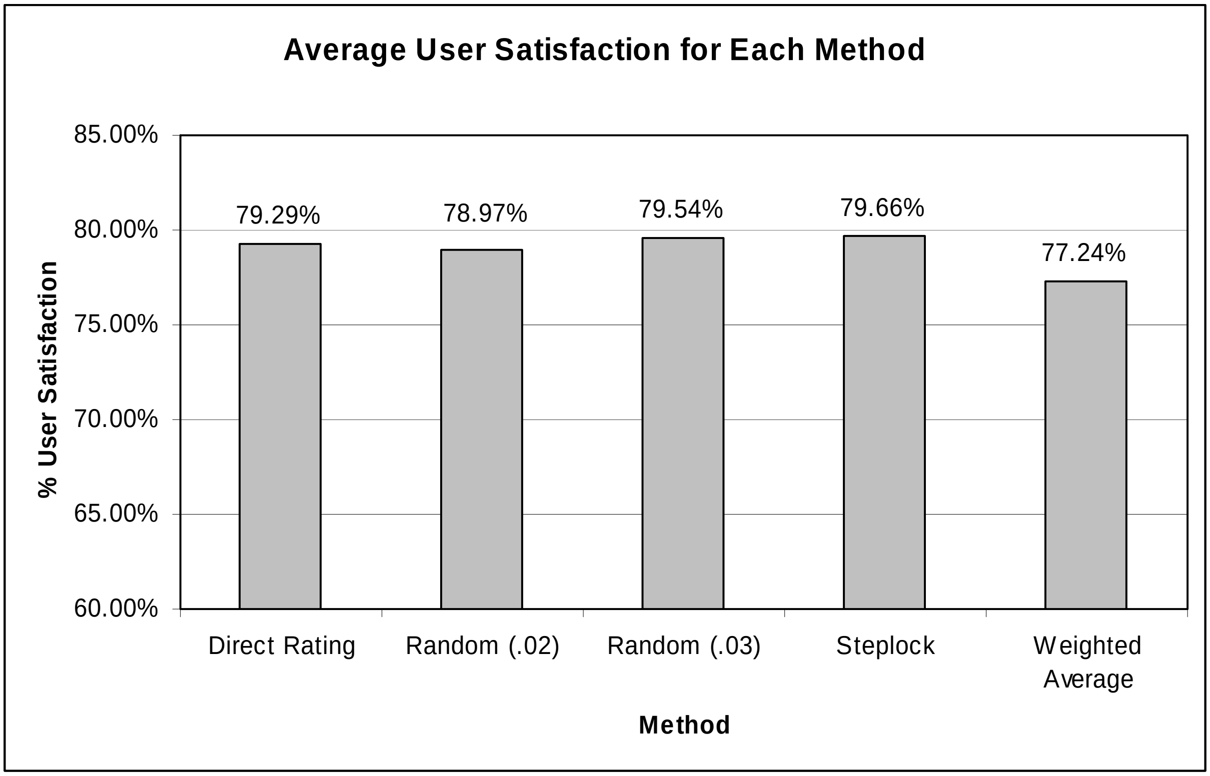
A t-test was run on this data. The results are summarized in Appendix C Table 11. No statistically significant difference in the number of votes needed to reach the final modifier was found between the Direct Rating, Steplock and Weighted Average Methods at an alpha value of 0.05. There was also no significant difference between the Random (.02) Method and the Random (.03) Methods. There was however a significant difference in the number of votes needed to reach the final modifier between the Random Methods and the Direct Rating, Steplock, and Weighted Average Methods.

In the Brown and overall categories, the Direct Rating Method produces the final modifier with the least user feedback. It is possible that this is the case because both Random Methods and the Steplock Method require several initial votes to move them from the midpoint of the possible range to the correct modifier. The Weighted Average Method, except in the case of the Green and Blue attributes, took more votes than the Direct Rating Method, but fewer than the other methods, to achieve its final membership value. This value was often different from the membership value arrived at by other methods, especially in the case of the Not and Slightly modifiers. Often the Slightly modifier was used by this method where other methods used the Not modifier. As suggested in [4] some of this is corrected by changing the upper bound of the Not modifier to be .10 instead of the .02 used in other methods.

Even with this change in the range, there were still cases where the modifier Slightly was chosen with the Weighted Average Method when other methods selected Not. It is possible that the range for Slightly and Not was set inappropriately and perhaps should be reworked for all the methods to produce more consistent results between the methods.

How well the image was categorized by the different methods was determined in the testing phase by a user satisfaction value. The user satisfaction rating only applies to the image overall, that is, a separate rating was not given to each color. Figure 12 shows the average user satisfaction for each method for each image.

Figure 12. Average User Satisfaction for Each Method



The average user satisfaction ranged from 77.24% to 79.66% with the Weighted Average Method having the lowest user satisfaction and the Steplock Method having the highest user satisfaction.

A t-test was done on this data and the results are summarized in Appendix C Table 12. At an alpha value of .05 there was no statistically significant difference between the percentages of user satisfaction for each method.

It is of interest that the user satisfaction ratings for each method were very similar although the methods did not all produce the same final modifier for each image. It would be expected that if the modifiers were different the percentage of satisfied users would be different. There are several possible reasons for this discrepancy. First, the way the questions were asked in the testing phase may not have elicited appropriate answers. Perhaps the term “satisfied” should have been replaced with “agree” to produce different results. Agreement may have a stronger meaning for some users than satisfied.

Additionally, many of the differences in the answers were in the Slightly and Not category. It is possible that these words are synonyms for each other for some colors. Another possibility is that range between these two modifiers is not set appropriately. It is possible that if Not is going to be its own category and not simply a cut off point, that it should be afforded a more sizeable portion of the modifier range, i.e. each modifier should get a quarter of the range. This change may help improve the overall performance of the Weighted Average Method, which in the case of the Slightly and Not modifiers, tended not to produce the same modifier as other methods.

Finally, although the Not modifier was added for each method, during the training phase, images in the Not category for a particular color were never presented to the user in the Random and Steplock Methods. In these methods images were shown to the user with the statements “People with <Slightly, Medium, Very> <Color> eyes.” The images in the Not category were shown in the Slightly category. Because the user was unaware that there was a Not category, it would be impossible for them to detect an image was in the Slightly category that should have been in the Not category and thus vote to move the image more towards the Not category.

6.2 Recommendations

The goal of this experiment was to determine which method of modifying the membership function produced the most accurate result with the smallest number of votes. Compared were the Direct Rating, Random with Step Size .02, Random with Step Size .03, Steplock and Weighted Average Methods. The Direct Rating Method produced the final modifier in the least number of votes. This method only differed from the highest percentage of user satisfaction by .41%, making it an obvious choice for use in determining a membership function.

This choice has two flaws, first the method did not have the highest satisfaction rating and second, the Direct Rating Method does not allow for changes after the initial membership function is converged. With the Direct Rating Method, after the initial training phase, there is no mechanism in place for user feedback to change the value of the membership function. Changing the value of the membership function after the training phase can be a valuable tool since the meaning of linguistic modifiers can change over time or with different user communities. It can also be used to correct misclassifications that occurred in the training phase.

The Steplock and Random Methods allow for user feedback as the database is used. If the system is queried for images with Slightly Blue Eyes, the user could give feedback as to whether they think individual images in that category are More Blue, Less Blue, or met the criteria. This feedback could be used to change the value of the membership function as was done during the training phase of this experiment. Additionally, the Steplock Method which has this feature, had the highest percentage of user satisfaction, meaning it achieves what the user feels is the best membership function for the image. The Steplock Method also used significantly fewer votes than the Random Methods and only slightly more than the Direct Rating and Weighted Average Methods.

It is recommended that two methods be used in conjunction to determine the value of the membership function. The Direct Rating Method should be used in the training phase to initially determine the value of the membership function, then the Steplock Method should be used as queries are asked of the system to handle ongoing user feedback. The step size in the Steplock Method should be appropriately small perhaps .01 or .005 to help prevent dramatic changes in the system should a data burst occur.

This method of determining the membership function handles the issue with the Direct Rating Method where two categories have nearly the same number of votes and thus a majority of users may not be satisfied with the result as (mentioned in section 3.3.1) by allowing further changes in the membership function as the system is used. Likewise, the issue of initialization of the Steplock Method mentioned in section 3.3.4 is handled by already training the system before using this method. The issue of throwing away votes with this method is not addressed; however, it is likely that over time this behavior could be desirable as the meaning of linguistic modifiers and the community or users changes.

CHAPTER 7 FUTURE WORK

Much work remains to be done on this subject. While this research aimed to answer several questions regarding constructing a membership function, several areas for further research were uncovered. First and foremost, the recommendation regarding how the membership function should be determined is untested and experiments should be run to ensure that this method is truly appropriate.

Further research should be done to determine where to set the ranges for the individual modifiers. The current scheme was set arbitrarily and it is possible that more appropriately set modifier ranges could promote convergence and raise the user satisfaction rate. It is possible that changing the ranges could have an effect on how the Weighted Average Method works and thus that method should be retested with new ranges.

Experiments should be run to test if the prototype should be changed for the Random and Steplock Methods to return images in the not category during the testing phase. This option could help decrease the number of images categorized as Not and Slightly by different methods. Additionally, Not and Slightly should be evaluated to see if they are synonyms.

The percentage of user satisfaction was very similar for all the methods of modifying the membership function. This raises several issues. An experiment should be run to determine if using a different word besides satisfied would produce a different and more accurate result. Tests should also be done to ensure that this metric is appropriate at all. A test could be conducted with images placed in categories using the methods tested and the images placed in categories randomly. Users would then be asked questions similar to the ones in the training phase for this experiment. In theory the user satisfaction should be lower for the randomly initialized images. If not it would suggest that user satisfaction determined this way is not an appropriate metric.

Throughout testing, users complained that the images were blurry and small. Testing should be done to determine if the prototype should be enhanced to allow for larger images or the ability to zoom in on the eyes. Additionally users should be questioned to determine which question format they prefer, that used in the Random and Steplock method or that used in the Direct Rating and Weighted Average Method. This can be used as a metric for determining an optimal method for modifying the membership function.

Different fuzzy attributes may produce different results with regards to these methods. It is possible that attributes such as nose length or face width could produce different results when these methods of modifying the membership function are used. Experiments should be run with different fuzzy attributes.

Literature Cited

Literature Cited

[1] Bilgic, T., Turksen I.B. (1999). "Measurement of Membership Functions: Theoretical and Empirical Work." *Handbook of Fuzzy Sets and Systems Vol 1, Fundamentals of Fuzzy Sets*, Chapter 3. Kluwer, pages 195-232.

[2] Boohtra, R. (2004). "Natural Language Interfaces: Comparing English Language Front End and English Query." Master's Thesis, Virginia Commonwealth University.

[3] Chameau, J. L. & Santamarina, J. C. (1987). "Membership functions part I: Comparing method of measurement. *International Journal of Approximate Reasoning 1*, pages 287-301.

[4] Cunningham, Joanne (2006). "Determining the Best Way to Process Community Opinions to Construct an Optimal Membership Function in a Fuzzy Database System." Proceedings of the 2006 ACM SE Conference, Vol. 1, pages 632-637.

[5] Dattarti, Smita & Joy, Karen (2004). "Implementing a Fuzzy Database and Querying System with Community Defined Membership Values." Directed Research Report, Virginia Commonwealth University.

[6] Dattari, Smita & Joy, Karen (2005). "Implementing a Fuzzy Relational Database Using Community Defined Membership Values." Proceedings of the 2005 ACM SE Conference, Vol. 1, pages 264-265.

[7] Hersh, H., Carmazza, A. & Briwnell, H.H. (1979). "Effects of context on fuzzy membership functions," in M. M. Gupta, A. Kendel, W. Bandler & J. B. Kiszka (eds), *Approximate Reasoning in Expert Systems*, North-Holland, Amsterdam, pages 33-46.

[8] Lee, Stephanie Scheibe (2005). "Fuzzy Membership Function Initial Values: Comparing Initialization Methods that Expedite Convergence." Master's Thesis, Virginia Commonwealth University.

[9] Leekwijck, Werner Van, Kerre E., Etienne. (1999). "Defuzzification: criteria and classification." Fuzzy Sets and Systems 108, pages 159-178.

[10] Mastros, James (2005). “Measuring Community Consensus in Facial Characterization Using Spatial Databases and Fuzzy Logic.” Master’s Thesis, Virginia Commonwealth University.

[11] McDermott, Jeff (2006). “Facial Feature Extraction for Face Characterization.” Proceedings of the 2006 ACM SE Conference Vol. 1, pages 750-751.

[12] Sanghi, Shweta (2006). “Determining Membership Function Values to Optimize Retrieval in Fuzzy Relational Database.” Proceedings of the 2006 ACM SE Conference Vol. 1, pages 537-542.

[13] Tashiro, H., Ohki, N., Nomura, T., Yokoyama T., Matsushita Y. (1993). “Managing Subjective Information in Fuzzy Database Systems.” Proceedings of the ACM Conference on Computer Science, pages 156-161.

[14] Zadeh, L.A. (1965). “Fuzzy Sets.” Information and Control 8, pages 338-353.

[15] Zadeh, L.A. (1972). “A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges.” Journal of Cybernetics 2, pages 4-34.

APPENDIX A

NUMBER OF VOTES NEEDED TO REACH FINAL VALUE FOR EACH METHOD

Table 5. Number of Votes Needed to Reach Final Modifier for the Direct Rating Method

Direct Rating				
Image ID	**Blue**	**Green**	**Brown**	**Overall**
1	4	11	1	11
2	4	1	2	4
3	43	77	3	77
8	2	1	1	2
9	4	1	1	4
10	35	24	1	35
11	72	1	1	72
12	1	13	1	13
13	1	1	2	2
14	85	1	1	85
15	1	86	1	86
16	16	4	1	16
17	4	2	1	4
18	1	1	2	2
19	5	30	1	30
21	1	1	2	2
23	1	1	2	2
24	2	1	1	2
31	34	1	1	34
32	2	1	10	10
33	8	1	2	8
34	9	2	1	9
35	41	1	1	41
36	23	1	1	23
37	47	1	1	47
38	4	1	1	4
40	1	1	1	1
Average	16.7037	9.88889	1.62963	23.1852

Table 6. Number of Votes Needed to Reach Final Modifier for the Random Method

Step Size .02

Random (.02)				
Image ID	**Blue**	**Green**	**Brown**	**Overall**
1	0	52	61	61
2	103	29	28	103
3	10	0	20	20
8	33	96	96	96
9	0	53	64	64
10	0	56	28	56
11	0	109	36	109
12	69	0	30	69
13	24	28	16	28
14	0	66	78	78
15	54	0	45	54
16	14	0	45	45
17	19	0	45	45
18	24	32	30	32
19	112	55	61	112
21	24	95	16	95
23	24	28	32	32
24	0	65	26	65
31	0	95	59	95
32	50	24	62	62
33	32	116	21	116
34	0	114	58	114
35	0	73	28	73
36	0	95	32	95
37	0	29	25	29
38	0	32	28	32
40	28	59	0	59
Average	22.963	51.8889	39.6296	68.1111

Table 7. Number of Votes Needed to Reach Final Modifier for the Random Method Step Size .03

Random (.03)				
Image ID	**Blue**	**Green**	**Brown**	**Overall**
1	0	80	75	80
2	81	61	76	81
3	62	0	44	62
8	74	80	37	80
9	18	72	76	76
10	73	77	78	78
11	0	69	75	75
12	79	0	81	81
13	55	74	15	74
14	85	80	78	85
15	83	58	79	83
16	7	0	79	79
17	84	0	76	84
18	67	70	15	70
19	5	64	78	78
21	69	73	9	73
23	74	70	15	74
24	39	67	75	75
31	0	80	76	80
32	74	71	15	74
33	82	80	53	82
34	0	80	75	80
35	0	73	75	75
36	0	80	75	80
37	0	80	77	80
38	0	76	75	76
40	84	77	19	84
Average	44.2593	62.6667	59.2963	77.7407

Table 8. Number of Votes Needed to Reach Final Modifier for the Steplock Method

Steplock				
Image ID	**Blue**	**Green**	**Brown**	**Overall**
1	0	24	24	24
2	20	0	18	20
3	20	0	25	25
8	18	24	43	43
9	15	60	22	60
10	0	68	18	68
11	15	20	22	22
12	20	0	28	28
13	20	18	7	20
14	58	29	26	58
15	20	0	18	20
16	19	62	22	62
17	66	0	26	66
18	16	18	21	21
19	5	29	22	29
21	18	18	15	18
23	18	18	21	21
24	58	22	18	58
31	0	24	22	24
32	18	18	39	39
33	42	30	22	42
34	0	34	22	34
35	0	20	20	20
36	0	22	20	22
37	24	20	18	24
38	0	20	22	22
40	18	23	0	23
Average	18.8148	23	21.5185	33.8148

Table 9. Number of Votes Needed to Reach Final Modifier for the Weighted Average Method

Weighted Average				
Image ID	**Blue**	**Green**	**Brown**	**Overall**
1	18	48	24	48
2	22	5	5	22
3	9	27	4	27
8	5	1	63	63
9	7	49	1	49
10	31	3	1	31
11	2	1	1	2
12	25	1	1	25
13	1	1	29	29
14	4	24	28	28
15	85	3	1	85
16	2	5	15	15
17	16	2	44	44
18	1	1	25	25
19	10	5	1	10
21	1	1	34	34
23	1	1	35	35
24	4	14	1	14
31	4	1	20	20
32	14	1	9	14
33	16	1	6	16
34	78	1	1	78
35	4	1	1	4
36	4	1	1	4
37	4	1	1	4
38	4	1	1	4
40	1	1	63	63
Average	13.8148	7.44444	15.4074	29.3704

APPENDIX B

PERCENTAGE OF USER SATISFACTION

Table 10. Percentage of Users Satisfied with Each Image

Percent Of User Satisfaction					
Image ID	**Direct**	**Rand_02**	**Rand_03**	**Steplock**	**Weighted**
1	77.48%	82.78%	82.78%	82.78%	66.00%
2	90.73%	90.73%	90.73%	86.67%	90.73%
3	76.67%	76.00%	76.00%	76.00%	77.33%
8	79.47%	79.47%	79.47%	79.47%	71.52%
9	70.86%	68.21%	69.33%	70.86%	69.33%
10	66.45%	68.21%	69.33%	66.45%	68.21%
11	77.48%	77.48%	77.48%	81.58%	77.48%
12	81.33%	77.33%	81.33%	77.33%	81.33%
13	86.75%	86.75%	86.75%	86.75%	86.75%
14	64.90%	64.67%	63.33%	63.33%	56.67%
15	82.00%	75.33%	82.00%	75.33%	75.33%
16	78.67%	78.00%	78.00%	78.00%	78.67%
17	60.93%	74.00%	72.00%	72.00%	48.00%
18	91.39%	91.39%	91.39%	91.39%	91.39%
19	76.00%	76.00%	76.00%	76.00%	76.00%
21	88.08%	88.08%	88.08%	88.08%	88.08%
23	90.07%	90.07%	90.07%	90.07%	90.07%
24	73.03%	73.51%	71.52%	73.51%	72.00%
31	80.13%	79.47%	79.47%	79.47%	80.13%
32	85.43%	85.43%	85.43%	85.43%	76.82%
33	61.59%	50.00%	56.67%	70.86%	64.24%
34	77.48%	77.48%	77.48%	77.48%	77.48%
35	85.53%	88.74%	88.74%	88.74%	88.74%
36	83.44%	83.44%	83.44%	83.44%	83.44%
37	84.77%	84.77%	84.77%	84.77%	84.77%
38	87.50%	87.42%	87.42%	87.42%	87.42%
40	82.78%	77.48%	78.67%	77.48%	77.48%

Percent Of User Satisfaction					
Image ID	**Direct**	**Rand_02**	**Rand_03**	**Steplock**	**Weighted**
Min	60.93%	50.00%	56.67%	63.33%	48.00%
Max	91.39%	91.39%	91.39%	91.39%	91.39%
Average	79.29%	78.97%	79.54%	79.66%	77.24%

APPENDIX C

t – Test Results

Table 11. t-Test Results for The Average Number of Votes Needed to Reach the Final Modifier for Each Method

Votes Needed to Reach Final Modifier, t-test Results			
Method 1	Method 2	t	Significance
Direct Rating	Random_02	5.736	Significant
Direct Rating	Random_03	10.113	Significant
Direct Rating	Steplock	1.710	Not Significant
Direct Rating	Weighted Average	0.898	Not Significant
Random_02	Random_03	1.651	Not Significant
Random_02	Steplock	5.195	Significant
Random_02	Weighted Average	5.356	Significant
Random_03	Steplock	13.040	Significant
Random_03	Weighted Average	10.793	Significant
Steplock	Weighted Average	0.816	Not Significant
df = 52, $\alpha = .05$, t = 2.403 for significance			

Table 12. t-Test Results for The Average User Satisfaction for Each Method

Percentage of Users Satisfied, t-test Results			
Method 1	Method 2	t	Significance
Direct Rating	Random_02	0.133	Not Significant
Direct Rating	Random_03	0.107	Not Significant
Direct Rating	Steplock	0.166	Not Significant
Direct Rating	Weighted Average	0.787	Not Significant
Random_02	Random_03	0.236	Not Significant
Random_02	Steplock	0.301	Not Significant
Random_02	Weighted Average	0.643	Not Significant
Random_03	Steplock	0.051	Not Significant
Random_03	Weighted Average	0.883	Not Significant
Steplock	Weighted Average	0.977	Not Significant
df = 52, α = .05, t = 2.403 for significance			

APPENDIX D

FINAL WEIGHTS, MODIFIERS, PERCENT OF USERS SATISFIED AND VOTES NEEDED TO REACH FINAL VALUE FOR EACH IMAGE AND METHOD

Table 13. Final Weight, Modifier, Percent of Users Satisfied, and Number of Votes to Reach Final Value for Each Image

		Final Weight			Final Modifier			Percent Satisfied	Votes Needed to Reach Final Value				
Image Id	Method	Blue	Green	Brown	Blue	Green	Brown	Total	Blue	Green	Brown	Average	Max
1	Direct	0.2	0	0	Slightly	Not	Not	77.48%	4	11	0	5	11
1	Random (.02)	0.52	0	0	Medium	Not	Not	82.78%	0	52	61	37.667	61
1	Random (.03)	0.5	0	0	Medium	Not	Not	82.78%	0	80	75	51.667	80
1	Steplock	0.5	0.02	0.02	Medium	Not	Not	82.78%	0	24	24	16	24
1	Weighted Avg	0.32	0.12	0.06	Slightly	Slightly	Not	66.00%	18	48	24	30	48
2	Direct	0	0.85	0	Not	Very	Not	90.73%	4	0	2	2	4
2	Random (.02)	0	0.86	0	Not	Very	Not	90.73%	103	29	28	53.333	103
2	Random (.03)	0	0.83	0	Not	Very	Not	90.73%	81	61	76	72.667	81
2	Steplock	0.02	0.62	0.02	Not	Medium	Not	86.67%	20	0	18	12.667	20
2	Weighted Avg	0.06	0.73	0.03	Not	Very	Not	90.73%	22	5	5	10.667	22
3	Direct	0	0.2	0	Not	Slightly	Not	76.67%	3	77	3	27.667	77
3	Random (.02)	0	0.52	0.12	Not	Medium	Slightly	76.00%	10	0	20	10	20
3	Random (.03)	0	0.5	0.15	Not	Medium	Slightly	76.00%	62	0	44	35.333	62
3	Steplock	0.02	0.41	0.08	Not	Medium	Slightly	76.00%	20	0	25	15	25
3	Weighted Avg	0.04	0.32	0.19	Not	Slightly	Slightly	77.33%	9	27	4	13.333	27
8	Direct	0	0	0.85	Not	Not	Very	79.47%	2	1	1	1.3333	2
8	Random (.02)	0	0	0.84	Not	Not	Very	79.47%	33	96	96	75	96
8	Random (.03)	0	0	0.83	Not	Not	Very	79.47%	74	80	37	63.667	80
8	Steplock	0.02	0.02	0.71	Not	Not	Very	79.47%	18	24	43	28.333	43
8	Weighted Avg	0.04	0.04	0.69	Not	Not	Medium	71.52%	5	1	63	23	63
9	Direct	0.2	0	0	Slightly	Not	Not	70.86%	4	1	1	2	4
9	Random (.02)	0.52	0	0	Medium	Not	Not	68.21%	0	53	64	39	64
9	Random (.03)	0.17	0.12	0	Slightly	Slightly	Not	69.33%	18	72	76	55.333	76
9	Steplock	0.23	0.02	0.02	Slightly	Not	Not	70.86%	15	60	22	32.333	60
9	Weighted Avg	0.16	0.12	0.06	Slightly	Slightly	Not	69.33%	7	49	1	19	49

		Final Weight			Final Modifier			Percent Satisfied	Votes Needed to Reach Final Value				
Image Id	Method	Blue	Green	Brown	Blue	Green	Brown	Total	Blue	Green	Brown	Average	Max
10	Direct	0.85	0	0	Very	Not	Not	66.45%	35	24	1	20	35
10	Random (.02)	0.5	0.18	0	Medium	Slightly	Not	68.21%	0	56	28	28	56
10	Random (.03)	0.17	0.09	0	Slightly	Slightly	Not	69.33%	73	77	78	76	78
10	Steplock	0.62	0.02	0.02	Medium	Not	Not	66.45%	0	68	18	28.667	68
10	Weighted Avg	0.46	0.18	0.01	Medium	Slightly	Not	68.21%	31	3	1	11.667	31
11	Direct	0.52	0	0	Medium	Not	Not	77.48%	72	1	1	24.667	72
11	Random (.02)	0.52	0	0	Medium	Not	Not	77.48%	0	109	36	48.333	109
11	Random (.03)	0.5	0	0	Medium	Not	Not	77.48%	0	69	75	48	75
11	Steplock	0.74	0.02	0.02	Very	Not	Not	81.58%	15	20	22	19	22
11	Weighted Avg	0.53	0.06	0.06	Medium	Not	Not	77.48%	2	1	1	1.3333	2
12	Direct	0	0.52	0	Not	Medium	Not	81.33%	1	13	1	5	13
12	Random (.02)	0	0.52	0.1	Not	Medium	Slightly	77.33%	69	0	30	33	69
12	Random (.03)	0	0.5	0	Not	Medium	Not	81.33%	79	0	81	53.333	81
12	Steplock	0.02	0.44	0.02	Not	Medium	Not	77.33%	20	0	28	16	28
12	Weighted Avg	0.06	0.46	0.09	Not	Medium	Not	81.33%	25	1	1	9	25
13	Direct	0	0	0.85	Not	Not	Very	86.75%	1	1	2	1.3333	2
13	Random (.02)	0	0	0.84	Not	Not	Very	86.75%	24	28	16	22.667	28
13	Random (.03)	0	0	0.83	Not	Not	Very	86.75%	55	74	15	48	74
13	Steplock	0.02	0.02	0.71	Not	Not	Very	86.75%	20	18	7	15	20
13	Weighted Avg	0.01	0.02	0.73	Not	Not	Very	86.75%	1	1	29	10.333	29
14	Direct	0.2	0	0	Slightly	Not	Not	64.90%	85	1	1	29	85
14	Random (.02)	0.38	0.12	0	Medium	Slightly	Not	64.67%	0	66	78	48	78
14	Random (.03)	0	0.06	0	Not	Slightly	Not	63.33%	85	80	78	81	85
14	Steplock	0.02	0.14	0.02	Not	Slightly	Not	63.33%	58	29	26	37.667	58
14	Weighted Avg	0.25	0.15	0.17	Slightly	Slightly	Slightly	56.67%	4	24	28	18.667	28

		Final Weight			Final Modifier			Percent Satisfied	Votes Needed to Reach Final Value				
Image Id	Method	Blue	Green	Brown	Blue	Green	Brown	Total	Blue	Green	Brown	Average	Max
15	Direct	0	0.2	0	Not	Slightly	Not	82.00%	1	86	1	29.333	86
15	Random (.02)	0	0.52	0	Not	Medium	Not	75.33%	54	0	45	33	54
15	Random (.03)	0	0.32	0	Not	Slightly	Not	82.00%	83	58	79	73.333	83
15	Steplock	0.02	0.41	0.02	Not	Medium	Not	75.33%	20	0	18	12.667	20
15	Weighted Avg	0.1	0.5	0.03	Not	Medium	Not	75.33%	85	3	1	29.667	85
16	Direct	0.2	0.2	0	Slightly	Slightly	Not	78.67%	16	4	1	7	16
16	Random (.02)	0.14	0.52	0	Slightly	Medium	Not	78.00%	14	0	45	19.667	45
16	Random (.03)	0.14	0.5	0	Slightly	Medium	Not	78.00%	7	0	79	28.667	79
16	Steplock	0.08	0.38	0.02	Slightly	Medium	Not	78.00%	19	62	22	34.333	62
16	Weighted Avg	0.27	0.29	0.05	Slightly	Slightly	Not	78.67%	2	5	15	7.3333	15
17	Direct	0	0	0	Not	Not	Not	60.93%	4	2	1	2.3333	4
17	Random (.02)	0.16	0.48	0	Slightly	Medium	Not	74.00%	19	0	45	21.333	45
17	Random (.03)	0	0.5	0	Not	Medium	Not	72.00%	84	0	76	53.333	84
17	Steplock	0.02	0.38	0.02	Not	Medium	Not	72.00%	66	0	26	30.667	66
17	Weighted Avg	0.14	0.27	0.14	Slightly	Slightly	Slightly	48.00%	16	2	44	20.667	44
18	Direct	0	0	0.85	Not	Not	Very	91.39%	1	1	2	1.3333	2
18	Random (.02)	0	0	0.86	Not	Not	Very	91.39%	24	32	30	28.667	32
18	Random (.03)	0	0	0.83	Not	Not	Very	91.39%	67	70	15	50.667	70
18	Steplock	0.02	0.02	0.71	Not	Not	Very	91.39%	16	18	21	18.333	21
18	Weighted Avg	0.02	0.02	0.74	Not	Not	Very	91.39%	1	1	25	9	25
19	Direct	0.2	0.2	0	Slightly	Slightly	Not	76.00%	5	30	1	12	30
19	Random (.02)	0.24	0.14	0	Slightly	Slightly	Not	76.00%	112	55	61	76	112
19	Random (.03)	0.2	0.2	0	Slightly	Slightly	Not	76.00%	5	64	78	49	78
19	Steplock	0.23	0.14	0.02	Slightly	Slightly	Not	76.00%	5	29	22	18.667	29
19	Weighted Avg	0.31	0.27	0.02	Slightly	Slightly	Not	76.00%	10	5	1	5.3333	10

		Final Weight			Final Modifier			Percent Satisfied	Votes Needed to Reach Final Value				
Image Id	Method	Blue	Green	Brown	Blue	Green	Brown	Total	Blue	Green	Brown	Average	Max
21	Direct	0	0	0.85	Not	Not	Very	88.08%	1	1	2	1.3333	2
21	Random (.02)	0	0	0.84	Not	Not	Very	88.08%	24	95	16	45	95
21	Random (.03)	0	0	0.83	Not	Not	Very	88.08%	69	73	9	50.333	73
21	Steplock	0.02	0.02	0.74	Not	Not	Very	88.08%	18	18	15	17	18
21	Weighted Avg	0.01	0.01	0.72	Not	Not	Very	88.08%	1	1	34	12	34
23	Direct	0	0	0.85	Not	Not	Very	90.07%	1	1	2	1.3333	2
23	Random (.02)	0	0	0.84	Not	Not	Very	90.07%	24	28	32	28	32
23	Random (.03)	0	0	0.83	Not	Not	Very	90.07%	74	70	15	53	74
23	Steplock	0.02	0.02	0.71	Not	Not	Very	90.07%	18	18	21	19	21
23	Weighted Avg	0.01	0.03	0.73	Not	Not	Very	90.07%	1	1	35	12.333	35
24	Direct	0.85	0	0	Very	Not	Not	73.03%	2	1	1	1.3333	2
24	Random (.02)	0.52	0	0	Medium	Not	Not	73.51%	0	65	26	30.333	65
24	Random (.03)	0.83	0.09	0	Very	Slightly	Not	71.52%	39	67	75	60.333	75
24	Steplock	0.41	0.02	0.02	Medium	Not	Not	73.51%	58	22	18	32.667	58
24	Weighted Avg	0.58	0.13	0.01	Medium	Slightly	Not	72.00%	4	14	1	6.3333	14
31	Direct	0.2	0	0	Slightly	Not	Not	80.13%	34	1	1	12	34
31	Random (.02)	0.42	0	0	Medium	Not	Not	79.47%	0	95	59	51.333	95
31	Random (.03)	0.5	0	0	Medium	Not	Not	79.47%	0	80	76	52	80
31	Steplock	0.56	0.02	0.02	Medium	Not	Not	79.47%	0	24	22	15.333	24
31	Weighted Avg	0.33	0.06	0.06	Slightly	Not	Not	80.13%	4	1	20	8.3333	20
32	Direct	0	0	0.85	Not	Not	Very	85.43%	2	1	0	1	2
32	Random (.02)	0	0	0.84	Not	Not	Very	85.43%	50	24	62	45.333	62
32	Random (.03)	0	0	0.83	Not	Not	Very	85.43%	74	71	15	53.333	74
32	Steplock	0.02	0.02	0.71	Not	Not	Very	85.43%	18	18	39	25	39
32	Weighted Avg	0.04	0.03	0.6	Not	Not	Medium	76.82%	14	1	9	8	14

		Final Weight			Final Modifier			Percent Satisfied	Votes Needed to Reach Final Value				
Image Id	Method	Blue	Green	Brown	Blue	Green	Brown	Total	Blue	Green	Brown	Average	Max
33	Direct	0	0	0	Not	Not	Not	61.59%	8	1	2	3.6667	8
33	Random (.02)	0.1	0.04	0.18	Slightly	Slightly	Slightly	50.00%	32	116	21	56.333	116
33	Random (.03)	0	0.06	0.5	Not	Slightly	Medium	56.67%	82	80	53	71.667	82
33	Steplock	0.02	0.02	0.53	Not	Not	Medium	70.86%	42	30	22	31.333	42
33	Weighted Avg	0.16	0.07	0.32	Slightly	Not	Slightly	64.24%	16	1	6	7.6667	16
34	Direct	0.52	0	0	Medium	Not	Not	77.48%	9	2	1	4	9
34	Random (.02)	0.48	0	0	Medium	Not	Not	77.48%	0	114	58	57.333	114
34	Random (.03)	0.5	0	0	Medium	Not	Not	77.48%	0	80	75	51.667	80
34	Steplock	0.5	0.02	0.02	Medium	Not	Not	77.48%	0	34	22	18.667	34
34	Weighted Avg	0.37	0.09	0.08	Medium	Not	Not	77.48%	78	1	1	26.667	78
35	Direct	0.85	0	0	Very	Not	Not	85.53%	41	1	1	14.333	41
35	Random (.02)	0.52	0	0	Medium	Not	Not	88.74%	0	73	28	33.667	73
35	Random (.03)	0.5	0	0	Medium	Not	Not	88.74%	0	73	75	49.333	75
35	Steplock	0.5	0.02	0.02	Medium	Not	Not	88.74%	0	20	20	13.333	20
35	Weighted Avg	0.63	0.04	0.03	Medium	Not	Not	88.74%	4	1	1	2	4
36	Direct	0.52	0	0	Medium	Not	Not	83.44%	23	1	1	8.3333	23
36	Random (.02)	0.52	0	0	Medium	Not	Not	83.44%	0	95	32	42.333	95
36	Random (.03)	0.5	0	0	Medium	Not	Not	83.44%	0	80	75	51.667	80
36	Steplock	0.44	0.02	0.02	Medium	Not	Not	83.44%	0	22	20	14	22
36	Weighted Avg	0.48	0.06	0.05	Medium	Not	Not	83.44%	4	1	1	2	4
37	Direct	0.52	0	0	Medium	Not	Not	84.77%	47	1	1	16.333	47
37	Random (.02)	0.52	0	0	Medium	Not	Not	84.77%	0	29	25	18	29
37	Random (.03)	0.5	0	0	Medium	Not	Not	84.77%	0	80	77	52.333	80
37	Steplock	0.38	0.02	0.02	Medium	Not	Not	84.77%	24	20	18	20.667	24
37	Weighted Avg	0.52	0.07	0.02	Medium	Not	Not	84.77%	4	1	1	2	4

		Final Weight			Final Modifier			Percent Satisfied	Votes Needed to Reach Final Value				
Image Id	Method	Blue	Green	Brown	Blue	Green	Brown	Total	Blue	Green	Brown	Average	Max
38	Direct	0.85	0	0	Very	Not	Not	87.50%	4	1	1	2	4
38	Random (.02)	0.52	0	0	Medium	Not	Not	87.42%	0	32	28	20	32
38	Random (.03)	0.56	0	0	Medium	Not	Not	87.42%	0	76	75	50.333	76
38	Steplock	0.5	0.02	0.02	Medium	Not	Not	87.42%	0	20	22	14	22
38	Weighted Avg	0.59	0.04	0.02	Medium	Not	Not	87.42%	4	1	1	2	4
40	Direct	0	0	0.85	Not	Not	Very	82.78%	1	1	1	1	1
40	Random (.02)	0	0.02	0.5	Not	Not	Medium	77.48%	28	59	0	29	59
40	Random (.03)	0	0.03	0.83	Not	Slightly	Very	78.67%	84	77	19	60	84
40	Steplock	0.02	0.02	0.59	Not	Not	Medium	77.48%	18	23	0	13.667	23
40	Weighted Avg	0.02	0.07	0.66	Not	Not	Medium	77.48%	1	1	63	21.667	63