Using T2FL Toolbox to Improve RAMSET’s Decision Making Fuzzy Model for Software Engineering Role Assignment

Luis G. Martinez, Juan R. Castro, Guillermo Licea, and Antonio Rodriguez-Diaz

Universidad Autónoma de Baja California
Calzada Tecnológico 14418, Tijuana, México 22300
{luisgm, jrcastro, glicea, ardiaz}@uabc.edu.mx
http://www.uabc.mx

Abstract. The choice of role selection for individuals in working team software developing projects is the basis of all Software Developing Process for organizations to be competitive in the industry. The present work is a follow up of RAMSET (Role Assignment Methodology for Software Engineering Teams) methodology that relates personality, abilities and software roles for integration of Software Engineering Teams, applying sociometric and psychometric techniques under a Fuzzy Approach. A type-1 FLS-based model was previously developed and implemented to manage uncertainty of personality test results; this paper proposes the use of type-2 fuzzy logic system so that more linguistic uncertainties can be handled and will lead to the creation of a better decision making model and continuous improvement of RAMSET methodology.

Keywords: Fuzzy Logic, Type-2 Fuzzy Logic, Uncertainty, Software Engineering.

1 Introduction

Effective use of psychometric instruments add value to organizations, these are used in selection and structured interview process to select more accurately people who will perform best in a role. Personality tests like Jung, Myers-Briggs, Big Five and projective tests like House-Tree-Person are used to know the sociopsychological characteristics and personality of individuals besides abilities for job placement and hiring and therefore to assign individuals to form a working team [1]. Personality tests are based on interpretation; therefore to tackle uncertainty we have found that using Fuzzy Logic has help us better define personality patterns thus better recommend best suited roles for performing in software engineering teams.

This paper is a study focused on the subjective Big Five Test used as a part of RAMSET’s methodology, a personality based methodology used in software project development case studies, first using a Takagi-Sugeno-Kang (TSK) Fuzzy Inference System (FIS) with an Adaptive Network Based Fuzzy Inference System (ANFIS) model approach, and then boarding the same case study using a Type-2 Fuzzy Logic (T2FL) Toolbox to compare uncertainty handling T2FL thus confirming reliability of the model.
The rest of the paper is organized as follows: section 2 is a brief background of personnel selection importance and related fuzzy logic approaches. Section 3 is a brief description of RAMSET methodology implemented. Section 4 defines our Big Five Fuzzy model, section 5 implementation of an ANFIS model, section 6 implementation of T2FL, section 7 displays results of the T2FL approach, concluding in section 8 with observations for discussion.

2 Background

Organizations' most important resource is the personnel who conforms them, specially qualified human resource. It is for that reason that selection of adequate personnel is the basis of Software Development Process in software engineering to be competitive in the market. In human resources selection of personnel is a decision making process where exists a range of uncertainty [2] to assure selection of adequate persons for the correct job. Researchers have propose different models and forms incorporating knowledge components such as data bases, rule systems, knowledge acquisition systems or dominion models to generate intelligent applications [3][4].

Psychologists in the industry who practice this area use work and candidate’s information to occupy the positions with the purpose of determining who is better qualified to carry out the position. Analysis of positions is the process to determine abilities, knowledge, capacities and required characteristics of personnel for a specific type of work. Based on the result of the analysis of positions, psychologists use adequate methods of selection to correlate the performance required with the offered work in specific.

Wei-Shen and Chung-Chian [5] propose a personnel selection tool based on viable fuzzy data mining techniques to assist in the search of an ideal candidate by means of reliable, effective and efficient information. Agreeing that future prediction of employee behavior must be the main point of personnel selection, trying to find the relationship between candidate attributes and expected behavior within the organization. Their tool assists a business administrator in finding qualified talent that fulfils necessary requirements more efficiently. Their method is based on three types of behavior: personnel must be induced to enter and remain in the organization; they must trustworthily carry out the required tasks of a position or specific role; and is required to be innovative and spontaneous in task accomplishment, able to go beyond the description of the entrusted role [6].

The approach of clarifying roles and responsibilities of team members, as well as the sense of job position expectation helps prevent effort duplication and diminish future conflicts. Psychological tests are fundamental part of decision making evaluation processes to establish differences between groups. When we developed our RAMSET methodology, we implemented different psychological tests, subjective tests like Myers-Briggs Type Indicator, Big Five and the projective Tree Test. With time and compilation of several cases we have found relationships between personality traits and software engineering roles assigned to people in working teams [7].
3 RAMSET Methodology

At the University of Baja California in Tijuana Mexico several case studies implemented RAMSET: a Role Assignment Methodology for Software Engineering Teams based on personality, in Software Engineering courses of our Computer Engineering Undergraduate Program. Uniqueness of this methodology is a combination of Sociometric techniques, Psychometrics and Role Theory in Software Engineering Development Projects, consisting of the next steps: (a) survey for abilities and skills, (b) implementation of Personality Tests, (c) execution of Personal Interviews, (d) implementation of Sociometric Technique, (e) assignment of Team Roles, (f) follow through of Team Role fulfillment.

RAMSET methodology has been described in previous publications, obtaining information on how to form teams [8] and description of personality patterns, specifically based on Tree Test [9], Jung [10] and Big Five [7] tests, thus we are on way to build a Decision Making Fuzzy Model for personnel selection with software support for each test.

In Software Engineering Development Teams each member can take a different role, for our case studies we adopted those defined by Tomayko [11]: architect: responsible for project creation, coordination and supervision; analyst: responsible for finding and following up on resources, requirement analysis and specifications; developer-programmer: responsible for implementation and code design; tester: responsible for tests and evaluation of the system; document specialist: responsible for compiling every document for evidence and defining documentation standards; image and presentation role: a representative in charge of selling and promoting the product.

As we can see Software Engineers have their typical fields of expertise like coordinating, designing, programming, testing, evaluating. The RAMSET methodology focus’ is to determine the best suited role for team performance relating personality test results with software engineering roles. We have worked with different Personality Tests all generating the dimension types or personality traits that display the behaviour of the individual member of the team. A fuzzy approach has given us valuable information for support in decision making assignment of roles, this paper specifically approaches Big Five Test using T2FL Toolbox to analyze uncertainty and test reliability of the test.

4 Big Five Fuzzy Model

Big Five personality tests claim to measure your intensities in relation to the Big Five factors. The structures of the tests require selecting options from multiple choice questionnaires. These big five personality tests equate your personality to your collective degrees of behavior in five factors. The Big Five factors are Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN, or CANOE if rearranged). The Neuroticism factor is sometimes referred to as Emotional Stability. And Openness factor sometimes is referred to as Intellect.

Openness (O) is a disposition to being imaginative, inventive, curious, unconventional and autonomous, and has an appreciation for art, emotion, adventure, unusual ideas, curiosity and a variety of experience. Conscientiousness (C) is comprised of two
related facets achievement and dependability. Has a tendency to show self-discipline, efficiency, organization, acts dutifully and aims for achievement, plans rather than behave spontaneously. Extraversion (E) represents tendency to be sociable, outgoing and assertive, possess positive energy, passion and excitement. Agreeableness (A) is a tendency to be trusting, friendly, compassionate, cooperative, compliant, caring and gentle. Neuroticism (N) represents a tendency to exhibit poor emotional adjustment and experience negative or unpleasant emotions easily, such as anxiety, insecurity, depression and hostility.

Bernstein et al. [12] have used the Big Five personality framework to predict the relationship between team member personalities and team effectiveness in the area of teamwork and decision making. They have found a differential pattern of relationships between the Big Five and the performance outcome, that is, traits such as conscientiousness, agreeableness, openness to experience, emotional stability and extraversion were positively related to teamwork, suggesting that individuals with these Big Five traits are more likely to be better team players. As team members are selected, organizations may want to consider the utility of these personality traits to improve the performance of the team, agreeing with Tupes and Christal [13].

Because of uncertainty of personality traits a fuzzy based approach is considered to provide an integrated quantity measure for abilities of software development personnel which incorporates all aspects of personality traits involved for role assignment. Fuzzy based approaches have been considered like Lathers [14] fuzzy model to evaluate suitability of Software Developers, also Ghasem-Aghaei and Orens [15] use of fuzzy logic to represent personality for human behavior simulation. Consequently encouraging engineering educators to make greater use of type theory when selecting and forming engineering design teams and delegating team roles, in benefit of achieving productivity and efficiency in team performance.

In psychology Bublakova et. al [16] have defined a fuzzy model for MMPI-2 (Minnesota Multiphasic Personality Inventory) to characterize personality features and psychic disorders, they implemented in MatLab Fuzzy Logic toolbox a support model to relate data obtained from the patient and prototypic profiles. Also Srivastava et. al. [17] have develop a method for method for analyzing and comparing group of students motivation using fuzzy logic. In robotics and human simulation Mustafa et al. [18] have used the latest psychological theories to design a complex dynamic system that reacts to any environment relying on fuzzy logic to simulate human emotional reaction based on five factor model (FFM) interpolating probabilities associated with personality traits and emotional response.

Classical set theory is based on binary logic: for each element we can say whether it belongs to a given set (1) or not (0), in fuzzy set theory not only an element can have a fully property or not at all, but it can also have partially property, quantified by a number from the interval [0, 1], thus giving us a tool to model vagueness in real life decisions. It allows us to describe mathematically linguistic values and linguistically defined rules [19]. Linguistic variables are used to manipulate imprecise qualitative and quantitative information; the linguistic variable is a variable whose values are not numbers but words or sentences in a natural or artificial language [20]. A linguistic variable is characterized by a quintuple (x, T(x), U, G, M), in which x stands for the name of the variable, T(x)
denotes the set of $x$ of fuzzy variable values, ranging over a universe of discourse $U$. $G$ is a syntactic rule for generating names of $x$, and $M$ is a semantic rule for associating each $x$ to its meaning being a subset of $U$.

The architecture of the Big Five Model proposed based on fuzzy logic is shown in figure 1, where each Big Five trait is an input linguistic variable. Openness (O) takes a label value of one (1), conscientiousness (C) a label value of two (2), extroversion (E) a value of three (3), agreeableness (A) a value of four (4) and neuroticism (N) a value of (5). Each variable has a numeric range value between 30 and 80. This model will calculate the degree of membership of personality traits to Role according to the initial values of the five input variables. Label values for output Role are (1) analyst, (2) architect, (3) developer-programmer, (4) documenter, (5) tester and (6) presenter.

![Diagram of Fuzzy Inference System Architecture for Big Five Test.](image)

**5 Big Five ANFIS Fuzzy Model**

Fuzzy Logic Toolbox software computes the membership function parameters that best allow the associated fuzzy inference system to track the given input/output data. The Fuzzy Logic Toolbox function that accomplishes this membership function parameter adjustment is called ANFIS. The acronym ANFIS derives its name from Adaptive Neuro-Fuzzy Inference System as defined by Jyh-Shing Roger Jang [21]. Using a given input/output data set, the toolbox function ANFIS constructs a Fuzzy Inference System (FIS) whose membership function parameters are tuned (adjusted) using either a back propagation algorithm alone or in combination with a least squares type of method.

The neuro-adaptive learning method works similarly to that of neural networks. This method has been applied to design intelligent systems for control [22][23][24], also for pattern recognition in areas like 3D object recognition[25], fingerprint matching [26] and human facial expression recognition[27]. Neuro-adaptive learning techniques provide a method for the fuzzy modeling procedure to "learn" information about a data set. Fuzzy Logic Toolbox software computes the membership function parameters that best allow the associated fuzzy inference system to track the given input/output data.
The modeling approach used by ANFIS is similar to many system identification techniques. First, you hypothesize a parameterized model structure (relating inputs to membership functions to rules to outputs to membership functions, and so on). Next, you collect input/output data in a form that will be usable by ANFIS for training. You can then use ANFIS to train the FIS model to emulate the training data presented to it by modifying the membership function parameters according to a chosen error criterion. In general, this type of modeling works well if the training data presented to ANFIS for training (estimating) membership function parameters is fully representative of the features of the data that the trained FIS is intended to model.

Taking advantage that neuro-adaptive learning techniques provide a method for "learning" information about a data set we also implemented an ANFIS model for our first-order Sugeno Fuzzy Model described earlier. The ANFIS under consideration has five variable inputs denoted by \( x = O, C, E, A, N \), with two Gaussian membership functions (B), a set of 32 rules and one output variable Role (R). Therefore a k-th rule can be expressed as:

\[
\text{IF } (x_1 \text{ is } B^{k}_{1}) \text{ AND } (x_2 \text{ is } B^{k}_{2}) \text{ AND } (x_3 \text{ is } B^{k}_{3}) \text{ AND } (x_4 \text{ is } B^{k}_{4}) \text{ AND } (x_5 \text{ is } B^{k}_{5}) \text{ THEN } R = f^{k}(x) \text{ where } f^{k} = p^{k}_{1} + p^{k}_{2} + p^{k}_{3} + p^{k}_{4} + p^{k}_{5}
\]

and membership functions are denoted by:

\[
\mu_{B^{k}_{i}} = \exp \left[ -\frac{1}{2} \left( \frac{x - m^{k}_{i}}{\sigma^{k}_{i}} \right)^2 \right]
\]  

(1)

where \( p_{i}^{k} \) are linear parameters, and \( B_{i}^{k} \) are Gaussian membership functions for \( k = 1, 2, 3, ..., 32 \) and \( i = 1, ..., 5 \).

The corresponding equivalent ANFIS architecture is as shown in Fig. 2. The entire system architecture consists of five layers, these are input, inputmf, rule, outputmf, output.

The set of rules obtained with ANFIS learning approach implemented in Matlab's commercial Fuzzy Logic Toolbox [28], help us simulate our case studies and has given us relationships between personality traits and software engineering Roles. Figure 3 shows this giving us what we have called Big Five Patterns (B5P) seeing significant differences between each trait-role relationship.

6 Big Five Type-2 Fuzzy Model

As we said uncertainty affects decision making and appears in a number of different forms resulting as some information deficiency, which may be incomplete, imprecise, fragmentary, not fully reliable, vague, contradictory or deficient in some other way [29]. Fuzzy reasoning handles much of this uncertainty, when something is uncertain it is difficult to determine its exact value, a type-1 fuzzy set makes more sense than using crisp sets [30]. However using an accurate membership function for something uncertain is not reasonable, so we need another type of fuzzy sets to handle these uncertainties called type-2 fuzzy sets [31].

Mendel [32] states that there are four sources of uncertainties in type-1 FLS: (1) meanings of words used in antecedents and consequents of rules can be uncertain
Fig. 2. ANFIS Architecture for Big Five Test.

Fig. 3. ANFIS Big Five Trait and Role Relationship.
(words meaning different things to different people), (2) Consequents may have a histogram of values associated with them, (3) measurements that activate a type-1 FLS may be noisy and therefore uncertain, (4) data used to tune type-1 FLS parameters may also be noisy. Type-1 fuzzy sets are not able to directly model such uncertainties because their membership functions are totally crisp. Type-2 fuzzy sets are able to model such uncertainties because their membership functions are themselves fuzzy, although type-2 fuzzy sets are more complicated their recent implementation is relevant [33][34][35][36][37].

In a type-1 FLS, output processing, called defuzzification, maps a type-1 fuzzy set into a number. Things are somewhat more complicated for an interval type-2 FLS, because to go from an interval type-2 fuzzy set to a number requires two steps[38]. The first step, called type-reduction, is where an interval type-2 fuzzy set is reduced to an interval-valued type-1 fuzzy set. The second step of Output Processing, which occurs after type-reduction, is still called defuzzification. Because a type-reduced set of an interval type-2 fuzzy set is always a finite interval of numbers, the defuzzified value is just the average of the two end-points of this interval. There can be two outputs to an interval type-2 FLS: crisp numerical values and the type-reduced set. The latter provides a measure of the uncertainties that have flowed through the interval type-2 FLS.

Therefore we are building an interval type-2 fuzzy neural network (IT2FNN) with help of a T2FL Toolbox [39]. One way to build IT2FNN is to fuzzyfy a conventional neural network. Each part of a neural network (activation function, weights, inputs and outputs) can be fuzzified.

The IT2FNN system is one kind of interval Takagi-Sugeno-Kang fuzzy inference system (IT2-TSK-FIS) in neural network structure. An IT2FNN is proposed, with TSK reasoning and processing elements called interval type-2 fuzzy neurons (IT2FN) for defining antecedents, and interval type-1 fuzzy neurons (IT1FN) for defining the consequents of rules.

Therefore a k-th rule can be expressed as:

\[ \text{IF } (x_1 \text{ is } \tilde{B}_1^k) \text{ AND } (x_2 \text{ is } \tilde{B}_2^k) \text{ AND } (x_3 \text{ is } \tilde{B}_3^k) \text{ AND } (x_4 \text{ is } \tilde{B}_4^k) \text{ AND } (x_5 \text{ is } \tilde{B}_5^k) \text{ THEN } R \text{ is } f^k(x) \]

where

\[ f^k(x) = [f^k_l, f^k_r] \]  \hspace{1cm} (2)

is defined by

\[ f^k_l = \sum_{i=1}^n c^k_i x_i + c_0^k - \sum_{i=1}^n s^k_i |x_i| - s_0^k; \quad \text{where } k = 1 \ldots 32. \]  \hspace{1cm} (3)

\[ f^k_r = \sum_{i=1}^n c^k_i x_i + c_0^k + \sum_{i=1}^n s^k_i |x_i| + s_0^k \]  \hspace{1cm} (4)

and the Gaussian interval type-2 membership function is denoted by:

\[ \mu_{F,k}(x) = \frac{\mu_{F_l,k}(x) \cdot \mu_{F_r,k}(x)}{\bar{\mu}_{F_l,k}(x) \cdot \bar{\mu}_{F_r,k}(x)} \]

defined by
\begin{align*}
1 \mu_{F_i^k}(x_i; [\sigma_i^k; 1 m_i^k]) &= \exp \left[ -\frac{1}{2} \left( \frac{x_i - 1 m_i^k}{\sigma_i^k} \right)^2 \right] \\
2 \mu_{F_i^k}(x_i; [\sigma_i^k; 2 m_i^k]) &= \exp \left[ -\frac{1}{2} \left( \frac{x_i - 2 m_i^k}{\sigma_i^k} \right)^2 \right] \\
\bar{\mu}_{F_i^k}(x) &= \begin{cases} 
1 \mu_{F_i^k}(x_i; [\sigma_i^k; 1 m_i^k]); & x_i < \frac{1}{4} m_i^k \\
1; & \frac{1}{4} m_i^k \leq x_i \leq \frac{3}{4} m_i^k \\
2 \mu_{F_i^k}(x_i; [\sigma_i^k; 2 m_i^k]); & x_i > \frac{3}{4} m_i^k 
\end{cases} \\
\mu_{F_i^k}(x) &= \begin{cases} 
2 \mu_{F_i^k}(x_i; [\sigma_i^k; 2 m_i^k]); & x_i \leq \frac{1}{4} m_i^k \\
1 \mu_{F_i^k}(x_i; [\sigma_i^k; 1 m_i^k]); & x_i > \frac{3}{4} m_i^k 
\end{cases}
\end{align*}

Being $\sigma$ standard deviation and $m$ mean of uncertainty where

\[ m_i^k \in [1 m_i^k; 2 m_i^k] \]

Figure 4 shows IT2-Membership functions of our IT2FL Big Five model.

![Fig. 4. IT2 Membership Functions of IT2FL-Big Five Model.](image)

## 7 Results

From 2007 through 2009 we have worked with software engineers operating in real projects, assigning 80 roles in software development teams. Of these 13 have been Analysts, 13 Architects, 17 Developers/Programmers, 14 Document specialists, 14 Tester and evaluators and 9 Image and presentation specialists.
This data was used for our decision making ANFIS model obtaining figure 3 showing relationships between each trait and their roles. On the horizontal axis the Input Trait ranges from 20 to 80, and in the vertical axis the Output Role result ranges from 1 through 6; where the numeric values for linguistic variable Roles were (1) Analyst, (2) Architect, (3) Developer-Programmer, (4) Documenter, (5) Tester and (6) Presenter. As discussed in previous work analysis of these graphics are that Openness trait (O) covers roles 2 thru 4, Conscientiousness trait (C) engulfs roles 4 thru 6, Extraversion trait (E) from 3 thru 6, Agreeableness trait (A) covers 2, 3 and 4, and Neuroticism trait (N) reaches all roles.

Results of trait means are 52.625 for O, 61.125 for C, 50.5125 for E, 51.25 for A and 62.525 for N. It is noted that data analysis range of trait means are from 50 to 62, so we will consider low degree around 30-40 and high degree around 70-80 based on standard deviation. With these results we can assert that a low degree of (E) is definitely recommended to place this person as a Developer-Programmer, this indicates a person that is highly Introverted has difficulty relating to others, although his high degree in trait (A) is an asset as he is very cooperative, trusting and compliant, attributes for a good programmer. Trait (N) is a significant trait as it involves a wide range of roles, those with low degree of (N) or better said with opposite trait a high degree of (ES) Emotional Stability is a quality of a leader presenting security, reassurance and self confidence; with a high degree of (N) a tester is recommended representing a person with insecurity, therefore with a high sense of wanting to check everything more than once.

After implementing the model with Type-2 membership functions we have obtained figure 4 showing the area of uncertainty for each trait. If a specific trait graphic shows a large area of uncertainty it means that this person can be assigned more different roles compared to a lower area on uncertainty. Our trait graphics show a small area of uncertainty, meaning that it has a higher level of certainty to assign fewer possible roles. If we relate O trait its range of roles are [1,4], C trait ranges [3,4], A trait [1,5], E and N traits cover all type of roles.

Trying to find some kind of variation between Type-1 fuzzy logic model and Type-2 fuzzy logic model there was none. Comparing figure 2 (Type-1) and figure 4 (Type-2) the difference is the uncertainty area, but it is minimum, meaning that the Big Five test has consistency, it is a test that has been refined thru the years so we can count on it’s results.

8 Conclusions

Fuzzy Logic has been permeating in different areas of science, use of FIS models in psychology and software engineering are creating new analytical methods and methodologies in research and development. Advantages of evaluating with decision making fuzzy models give the process a robust and consistent degree of subjectivity to the evaluator, in our case study it provides a reliable personnel selection methodology to assign roles in software engineering teams.

Empirical results obtained are confirming our valid approach using fuzzy models as software support in role assignment by RAMSET methodology. The use of T2FL
Toolbox confirms reliability of tests implemented, future work with all personality tests will assure consistency and reliability of our methodology.

Implementation of ANFIS models is a highly powerful tool to improve Data Base Rules arisen from this study; combination of different personality test FIS models will create a computer aided software tool invaluable for decision making in assignment of software engineering roles.

References