

Apply Different Fuzzy Integrals in Unit Selection Problem of Real Time Strategy Game

Y.J. Li, Peter H. F. Ng, H. B. Wang, S. C. K. Shiu
Department of Computing
Hong Kong Polytechnic University
Hong Kong, China
(csyjli, cshfng, cshbwang, cscskshiu)@comp.polyu.edu.hk

Y. Li
College of Mathematics & Computer Science
Hebei University
Baoding, China
polyuliyen@yahoo.com.cn

Abstract—Choquet Integral (CI), which is known as a fuzzy measure-based technique, has been a general aggregation tool for multi-criteria decision making problem. In this paper, we apply Choquet Integral to unit selection problem in Real Time Strategy (RTS) game. In addition, three new fuzzy integrals named Mean-based Fuzzy Integral (Me-based FI), Max-based Fuzzy Integral (Ma-based FI), and Order-based Fuzzy Integral (Or-based FI) are developed, which relax the monotonicity requirement of the traditional fuzzy measures and consider different properties of game play. We compare the performance of Choquet Integral and the new proposed ones on this practical application with highly non-monotonic data. Experiments show that the proposed new fuzzy integrals achieved better learning performance and testing result.

Keywords—Real time strategy game; feature interaction; fuzzy integral; fuzzy measure; Warcraft III

I. INTRODUCTION

Sugeno [1] (1974) proposed the concepts of fuzzy measure and fuzzy integral which are working by finding the disordered means among cases. Suppose we have a dataset X , and A is a subset of it. As an extension of the probability measure, the fuzzy measure of A , expressed as $\mu(A)$, demonstrates the degree of belief or confidence of its elements. Fuzzy integrals cover a wide range of aggregation criteria and able to represent the interactions among variables. In addition, Choquet Integral is a powerful nonlinear aggregation tool. By considering non-additive set function, it has been widely used in information fusion [2] and data mining [3]. Mostly, the adopted non-additive set function is referred to as the class of fuzzy measures which are constrained by the monotonic property. Although Murofushi [4] has proved that the Choquet Integral is meaningful in the case of non-monotonic situation, there are few literatures in real application.

Real time strategy (RTS) game is a genre of computer war-games which does not process incrementally in turn. Players need to control the use of limited resource, create units and buildings, and lead their armies to attack their opponents. This kind of interactive game provides a rich environment for incremental research on human-level Artificial Intelligence (AI), especially real-time planning [6] and decision making under uncertainty [7]. The core of strategy planning is to build up the army with appropriate unit types which can gain massive destroy power against enemy's army in a short time. On the other hand, the army is taking great risk of being easily

defeated by other kinds of units [16]. Thus, understand the effect of unit combination in this complicated game play design is a different problem and it already becomes one of the major challenges to game players. Furthermore, there are feature interactions existing among different unit types ranging from redundancy to synergy [17]. In this situation, the effectiveness of unit combinations cannot be simply computed by weighted average. This leads to a more difficult task of dealing with non-additive properties in game. In this paper, we try to apply fuzzy integrals in unit selection problem of RTS game.

The rest of the paper is organized as follows. In section II, we review the fuzzy measure and Choquet Integral. Section III explains our methodology of fuzzy measure learning with different integrals. Section IV compares the performance of Choquet Integral and our new developed fuzzy integrals on game problem. Finally, conclusions are drawn in Section V.

II. FUZZY MEASURE AND CHOQUET INTEGRAL

To increase the modeling ability, the original non-additive fuzzy measure is preferred. Some experiments are performed by using Sugeno- λ and K-additive fuzzy measure, but unfortunately, both of them achieved poor results with high error. One possible reason is the sense of monotonic assumption. Building a balanced army with many kinds of unit types is improper in game environment.

In this section we review the concepts of fuzzy measure and Choquet Integral. The concept of non-monotonicity and some related properties of the Choquet Integral will also be given.

A. Fuzzy Measure

Additive property is the main characteristic of classical measures. This is an effective and convenient nature, but is often too rigid in many practical applications. Fuzzy Measure extends the application area by replacing the additive property with weaker condition as monotonicity.

Definition 1: A fuzzy measure on a measurable space (X, F) is a real-valued set function $\mu : F \rightarrow (-\infty, +\infty)$ satisfying:

$$(MM1) \mu(\emptyset) = 0,$$

$$(MM2) \mu(A) \geq \mu(B) \text{ whenever } A \in F, B \in F, A \subseteq B.$$

In case of emphasizing the monotonicity, we can also call it a monotonic fuzzy measure [4].

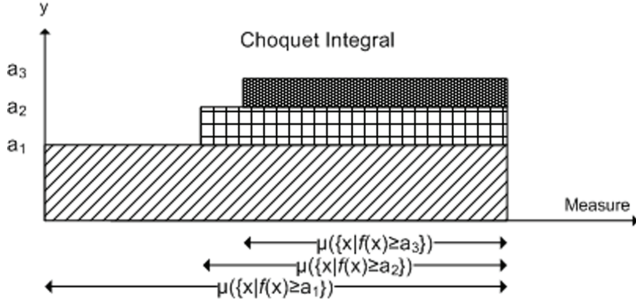


Figure 1. Choquet integral. μ is the fuzzy measure of function $f(x)$

Definition 2: A non-monotonic fuzzy measure on (X, F) is a real-valued set function $\mu : F \rightarrow (-\infty, +\infty)$ satisfying $\mu(\emptyset) = 0$.

B. Choquet Integral

Choquet Integral (Figure 1), as a classical fuzzy measure and an expansion of Lebesgue Integral, is defined by **Definition 3**. Over decades, it has gained considerable attention and success in multi-criteria decision making and data mining.

Definition 3: Suppose a fuzzy measure μ on X . The discrete Choquet integral of a function $f : X \rightarrow \mathbb{R}^+$ can be written as:

$$(c) \int f(x) \circ \mu(X) = \sum_{i=1}^n [f(x_i) - f(x_{i-1})] \cdot \mu\{x | f(x) \geq f(x_i)\} \quad (1)$$

Where x_i is permutation on X and $f(x_0) = 0$.

Choquet Integral has the following 3 basic properties:

- 1) $(c) \int 1_A d\mu = \mu(A)$
- 2) If $f < g$ then $(c) \int f d\mu \leq (c) \int g d\mu$
- 3) If a is non-negative real number and b is a real number, then

$$(c) \int (af + b) d\mu = a(c) \int f d\mu + b\mu(X)$$

As Murofushi introduced in [4], non-monotonic measure occurs when resources are limited. The basic properties of the Choquet integral are still useful for a non-monotonic fuzzy measure. RTS game provides a very similar environment with limited time and resource. If player wastes his resource in developing overmuch types of unit, the enemy will get opportunities to destroy his resources.

III. PROPOSED METHODOLOGY

RTS game provides a complicated environment which consists hundreds of buildings and units with different properties [5]. The various characters lead to various combinations. Even the most expensive and powerful units can be easily defeated by cheap units. In this setting, pro-gamer prefers to develop an unbalanced army with only several unit

types [15]. Different combination performs different destroy power to certain kinds of unit types and buildings. Furthermore, the combinations are usually affected by a third dependent variable. By tracing the players' historical action, we find that there are substantial feature interactions among units. For example, melee unit often cooperates with magic and rang unit. With this positive interaction, the model is considered as super additive, i.e., $f(x+y) \geq f(x) + f(y)$. When the interaction is negative, the model is considered as sub additive, i.e., $f(x+y) \leq f(x) + f(y)$.

Quantitative interactions among groups of data can be detected by Analysis of Variance (ANOVA) test [8]. If the calculated p-value is smaller than a certain level (usually 5%), we can acquire a strong confidence to confirm that there are interactions among these testing variables. In this section, we describe how to learning fuzzy measure with different integrals for RTS game.

A. Strategy Definition in RTS Games

Strategy is defined as arranging the army with appropriate unit selection which can gain massive destroy power against enemy's unit. The knowledge was extracted from replay file which is provided by game system of Warcraft III. Each replay file consists of sequential records of players' actions which are designed for reviewing the game battle. We develop a C# program to decrypt and extract the features from replay file. Each battle represents a strategy case and has been divided into three parts as goal, situation, and scores.

$$\text{Strategy Case} = \{\text{Goal, Situation, Scores}\}$$

"Goal" is the creation of a suitable army mixed with certain proportion, e.g., 10% peasants, 50% footman and 40% rifleman.

"Situation" is used to describe what circumstances the player is dealing with. In this paper we define it as enemy's unit combination. Cases will be clustered into different groups with different races and unit combinations.

"Score" is the point obtained from each battle. It can be regarded as the outcome evaluation for each in-game player. Higher performance will represent better unit combination. We use these scores in learning the fuzzy measure.

B. Learning Fuzzy Measure by GA

Genetic Algorithm (GA) is a stochastic search technique with a parallel random search mechanism which aims to achieve a global minimum on error space. We use it to learn the non-additive fuzzy measure in our application problem [9][10]. The flowchart is shown as in Figure 2.

In Warcraft III, each player will get several scores in each battle as the performance evaluation. The higher score denotes the higher chance of winning. We define our score, i.e., *Score* as the differences between player and enemy.

$$\text{Score}(x) = \text{Score}_{\text{WarcraftIII, player}} - \text{Score}_{\text{WarcraftIII, enemy}} \quad (2)$$

Fuzzy integrals are used to calculate the estimated scores which combine unit production statistic (i.e., $f(x)$) and non-monotonic fuzzy measure. Each fuzzy measure $\mu(x)$ describes

an interaction of the corresponding unit combination. The $Score(x)$ will be normalized between 0 to 1 and compare with the estimated score.

$$Score_{Different}(x) = |Score_{Estimate}[f(x)] - Score(x)| \quad (3)$$

Our fitness function is designed to obtain the average differences between real scores and estimated scores in all the training replays. We apply Wang's fitness function as shown in (5).

$$Fitness\ value = -\frac{1}{1+e} \text{ where } e = \left(\frac{1}{N} \sum_{i=1}^N score_{Different}^2(i) \right)^{\frac{1}{2}} \quad (4)$$

In our GA algorithm, each chromosome consists 2^{n-1} fuzzy measures of $\mu(X)$ including the entire unit combinations i.e.: $\mu(x_1), \mu(x_2), \dots, \mu(x_1, x_2), \dots, \mu(x_1, x_2, \dots, x_n)$. $x_1, x_2, \dots, x_n$ represent n kinds of unit types. Each fuzzy measure is first corrected to seven decimal places and then converted into binary form. A number of chromosomes will be initialized randomly at first. The learning process is repeated until the fitness value is stable or exceeds the maximum generation. After that, a reasonable fuzzy measure $\mu(X)$ for training replays is obtained. Population is set as 50, maximum generation is set as 100 and mutation rate is set as 0.01. The detail of the learning process are introduced in [11] [12].

C. Different Fuzzy Integrals

As seen from Figure 2, we define the function $f(x)$ which represents a player unit statistic of one replay:

$$f(x) = f(\text{Unit Statistic}_{\text{Replay } i, \text{Player}}).$$

All the unit statistics are normalized and presented by proportion. E.g.: 4 footmen, 6 riflemen and 10 priests will be normalized into $f(x_1)=0.2, f(x_2)=0.3$ and $f(x_3)=0.5$ respectively. This definition is useful in all the fuzzy integrals proposed in this paper.

We use Choquet Integral to calculate the estimated score as an example. Suppose there are 7 unit types, i.e.: $n=7$ and the unit statistic of a replay is $f(X)=\{x_1=0.2, x_2=0.4, x_3=0, x_4=0.3, x_5=0, x_6=0.1, x_7=0\}$. Then, ascending order sorting is performed with respect to the proportion, i.e.: $a_1=x_3=0, a_2=x_5=0, a_3=x_7=0, a_4=x_6=0.1, a_5=x_1=0.2, a_6=x_4=0.3, a_7=x_2=0.4$ where $a_1 \leq a_2 \leq \dots \leq a_7$. Given a fuzzy measure as $\mu(X)=\{\dots, \mu(x_6)=0.4, \mu(x_4, x_6)=0.7, \mu(x_1, x_4, x_6)=0.8, \mu(x_1, x_2, x_4, x_6)=0.2, \dots\}$, the estimated score is computed by:

$$\begin{aligned} & (c) \int f(x) \circ \mu(X) \\ &= \sum_{i=1}^7 (a_i - a_{i-1}) \cdot \mu(\{x \mid f(x) \geq a_i\}) \\ &= ((a_1 - a_0) \cdot \mu(x \mid f(x) \geq a_1)) + ((a_2 - a_1) \cdot \mu(x \mid f(x) \geq a_2)) + \dots \\ &= 0 + ((a_4 - a_3) \cdot \mu(x \mid f(x) \geq a_4)) + ((a_5 - a_4) \cdot \mu(x \mid f(x) \geq a_5)) + \dots \\ &= 0.1 \cdot \mu(x_1, x_2, x_4, x_6) + 0.1 \cdot \mu(x_1, x_4, x_6) + 0.1 \cdot \mu(x_4, x_6) + 0.1 \cdot \mu(x_6) \\ &= 0.1 \times 0.2 + 0.1 \times 0.8 + 0.1 \times 0.7 + 0.1 \times 0.4 \\ &= 0.21 \end{aligned}$$

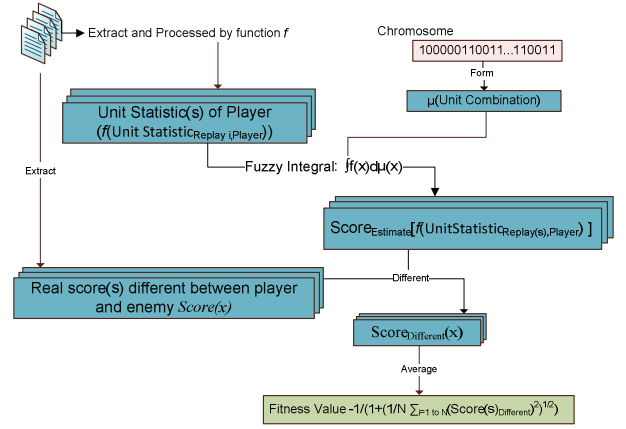


Figure 2. Fitness calculation of one chrommosom

Choquet integral is proved to work in our non-monotonic situation. Unfortunately, according to its fuzzy measure selection, the unit with least production has interaction with all the other units (i.e., basic unit) and the units with most production (i.e., powerful unit) have no interaction with others. This selection cannot well support the Warcraft III game play design. Thus, we develop three new fuzzy integrals based on different considerations. They are named as Max-based Fuzzy Integral (Ma-based FI), Mean-based Fuzzy Integral (Me-based FI) and Order-based Fuzzy Integral (Or-based FI).

1) Max-based Fuzzy Integral

Definition 4: Suppose a fuzzy measure μ on X . The Max-based Fuzzy Integral of a function $f: X \rightarrow \mathbb{R}^+$ can be written as:

$$(M) \int f(x) \circ \mu(X) = \sum_{i=1}^n x_i \times \max(\mu(S_i)) \quad (5)$$

Where $x_i \in S_i$ and $\forall x \in S_i, x \neq 0$

Ma-based FI is developed with the thinking of winner takes all. We just consider the most powerful unit combinations which involve the current unit. This setting may lead the fuzzy measure learning process focuses on the important options.

2) Mean-based Fuzzy Integral

Definition 5: Suppose a fuzzy measure μ on X . The Mean-based Fuzzy Integral of a function $f: X \rightarrow \mathbb{R}^+$ can be written as:

$$(m) \int f(x) \circ \mu(X) = \sum_{i=1}^n x_i \times \left(\frac{1}{m_i} \sum_{j=1}^{m_i} \mu(S_{ij}) \right) \quad (6)$$

Where $x_i \in S_{ij}$ and $\forall x \in S_{ij}, x \neq 0$, n is number of unit type, m is the number of set which include x_i

Me-based FI is developed to release the problem of interactions selection. All the fuzzy measures which involve the current unit type will be selected and an average is computed.

3) Order-based Fuzzy Integral

Players may need to develop technology and buildings to unlock advanced unit where the resource weighting in advanced unit is higher and more powerful. According to some data analysis, we found that advanced unit usually dominates the proportion of unit statistic. We infer that most players focus on building their army with advanced units. The advanced units should be considered as having more cooperation with other units. Thus, the largest $f(x)$ is combined with the fuzzy measure with most of the units, and the unit type with least production is considered as having no interaction with others.

As Figure.3 shows, players will create labor at the very beginning to exploit resource. Then some basic units will be created for attack or defense. Through some development on building and technology, different kinds of advanced unit are unlocked. Constrained by resource, player will concentrate on certain unit types to build up his army. Proportion of unit combination will also be dominated by the selected unit types in order to maximize the destroy power.

Or-based FI is developed by consider the unit production sequence. Each unit type calculates the interaction with the one who has less production than it. The definition of Or-based FI is shown as follows.

Definition 6: Suppose a fuzzy measure μ on X . The Or-based Fuzzy Integral of a function $f : X \rightarrow \mathbb{R}^+$ can be written as

$$(o) \int f(x) \circ \mu(X) = \sum_{i=1}^n a_i \cdot \mu(x \mid 0 < f(x) \leq a_i) \quad (7)$$

Where $a_1 \geq \dots \geq a_{n-1} \geq a_n > 0$

Different from CI, descending order sorting is performed among unit types in Or-based FI. Adopting the same example used in CI, the estimated score is computed as follows.

$$\begin{aligned} (r) \int f(x) \circ \mu(X) &= \sum_{i=1}^4 (a_i) \cdot \mu(\{x \mid 0 < f(x) \leq a_i\}) \\ &= a_1 (\mu(a_1, a_2, a_3, \dots, a_7)) + a_2 (\mu(a_2, a_3, \dots, a_7)) + \dots + a_4 (\mu(a_7)) \\ &= x_2 (\mu(x_1, x_2, x_4, x_6)) + x_4 (\mu(x_1, x_4, x_6)) + x_1 (\mu(x_1, x_6)) + x_6 (\mu(x_6)) \\ &= 0.4 \times 0.2 + 0.3 \times 0.8 + 0.2 \times 0.7 + 0.1 \times 0.4 \\ &= 0.5 \end{aligned}$$

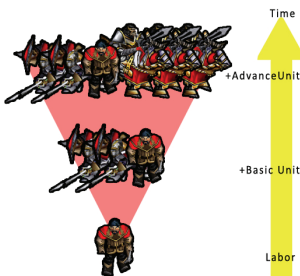


Figure 3. Unit production sequence in Warcraft III

4) Set Selection for Different Fuzzy Integrals

The main difference of these four fuzzy integrals is the set selection. Each $f(x)$ will select different fuzzy sets for different fuzzy integrals. According to Wang's [13], the fuzzy measure selection function could be generated by the following equations.

$$K_i = \left\{ k : \frac{i}{2^k} - \left\lfloor \frac{i}{2^k} \right\rfloor \geq 0.5, 1 \leq k \leq n \right\}$$

$$\overline{K}_i = \{1, 2, \dots, n\} - K_i$$

where $i=1, 2, \dots, 2^n-1$ and $k=1, 2, \dots, n$, n is the number of feature $\lfloor a \rfloor$ denotes the integer part for a non-negative number a .

Let δ_i be the set selection operator as defined in (8). Equation (9)-(12) expresses the fuzzy measure selection for different fuzzy integrals: Choquet Integral (10), Or-based- FI (11), Me-based FI (12), and Ma-based FI (13).

$$\delta_i = \max(\min_{k \in K_i} f_k - \max_{k \in \overline{K}_i} f_k, 0) \quad (8)$$

$$\delta_i = \begin{cases} \min_{k \in K_i} f_k - \max_{k \in \overline{K}_i} f_k, & \min_{k \in K_i} f_k > \max_{k \in \overline{K}_i} f_k \\ 0, & \min_{k \in K_i} f_k \leq \max_{k \in \overline{K}_i} f_k \end{cases} \quad (9)$$

Assume $\min_{k \in \emptyset} f_k = \max_{k \in \emptyset} f_k = 0$

$$\delta_i = \begin{cases} \max_{k \in K_i} f_k, & \max_{k \in K_i} f_k < \min_{k \in \overline{K}_i} f_k \\ 0, & \max_{k \in K_i} f_k \geq \min_{k \in \overline{K}_i} f_k \end{cases} \quad (10)$$

$$\delta_i = \text{avg}_{k \in K_i} f_k \quad (11)$$

$$\delta_i = \max_{k \in K_i} f_k \quad (12)$$

The sorting order makes Choquet Integral cannot directly calculate the maximum value. But this mechanism can be applied into Or-based FI. The selected maximum value is then combined with the corresponding unit statistic to calculate the estimated score. The set selection will not involve the assumption of $\min_{k \in \emptyset} f_k = \max_{k \in \emptyset} f_k = 0$, while empty set property of $\min_{k \in \emptyset} f_k = \infty$ is available.

IV. EXPERIMENT

A. Data Collection and Preprocessing

We selected Warcraft III as our research platform. It is a famous RTS game released in 2003 with the award of "Game of the Year" from more than six game publications. It has been sold over 7 million copies and there are concurrently over 200,000 players playing 6,000 battles online.

2,649 replay files of professional one-versus-one competitions from the internet are collected. They consist of two parts of data as shown in TABLE I. The header contains

some demographic data and the other part is composed of different action blocks. Each block is encrypted and stored players' actions in every 250 milliseconds' interval. A detailed example of replay file is shown in TABLE II.

The C# program extracts three kinds of data from these replays: (i) player unit production statistics, (ii) enemy unit type, and (iii) performance scores. Five data sets, as described in TABLE III, with largest amount of cases are prepared for our research. Each data set consist the cases of one race encountering its enemies which have a certain unit combination. Beside two less attended unit types, Figure 4 shows that the unit production follows a normal distribution.

TABLE I. DATA IN WARCRAFT III REPLAY

Element	
Header	Player Record, Game Name, Map Settings, Map Record, Map & Creator Name, Player Count, Game Type, Language ID, Player List, Game Start Record, Slot Record, Random Seed,
Actions Block	Player ID, Action ID, Action Arguments

TABLE II. SAMPLE OF WARCRAFT III REPLAY

Time	Action
00:00:02:002	Player 1 train 1 Peasant
00:00:02:253	Player 1 select 5 [Peasant]
00:00:02:503	Player 1 Right Click with 0x58a8 at(7296,2432)
00:00:15:241	Player 2 build Altar of Darkness at(-1376,6240)

Player 1 produces 1 Peasants, then select 5 Peasants and move to (X: 7296, Y: 2432). Player 2 builds Altar of Darkness at(X: -1376, Y: 6240)

TABLE III. NATURE OF TESTING DATA SET

Data Set	1	2	3	4	5
Player race	Undead	Undead	Orc	Orc	Elf
Enemy unit	Fm, P	Dr, Fm, P	Pr, So, Sb, P	Fm, P	Fm, P
No. of case	162	65	1004	549	869
No. of Combination	23	16	60	31	31

Fm – Footman, P – Peasant, Dr, Dragon rider, Pr – Priest, So – Sorceress, Sb – Spell break

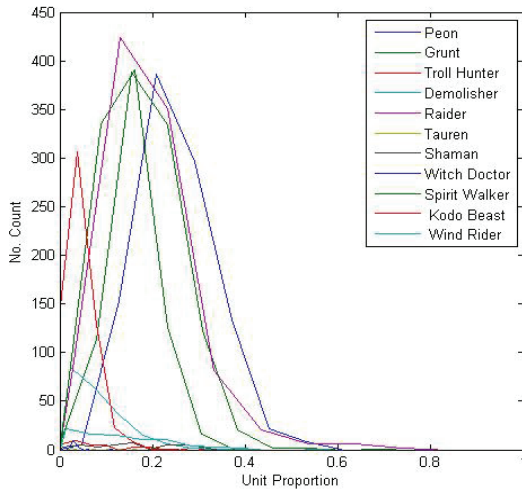


Figure 4. Unit production sequence in Warcraft III

B. Experimental Comparison of Different Fuzzy Integrals

In this part, we compare the performance of the four related fuzzy integrals. All the simulations are performed in MATLAB. The machine used is an Intel premium 4 2. 3GHZ with 2GB Ram PC. 70% cases in each dataset are used for training and the rest 30% are for testing.

Classical weight average (WA) method is also considered since we want to survey whether considering interaction can attain better result in our Warcraft III game data. The setting is similar as Yeung's [14] approach. Each unit type is given a weighting w and the unit combination power is calculated as the summation of the corresponding unit type's weighting. The learning process applies genetic algorithm with the same parameters settings as described in section 4.

$$f(X) = \frac{1}{n} \sum_{i=1}^n w_i x_i \quad (13)$$

Error is computed as the absolute difference between the estimated score (calculated by fuzzy integrals) and the real

TABLE IV. TRAINING ERROR IN DIFFERENT FUZZY INTEGRAL

Data Cluster	1	2	3	4	5
WA	0.512	0.576	0.799	0.497	0.463
CI	0.194	0.201	0.485	0.195	0.059
Ma-based FI	0.160	0.245	0.048	0.140	0.197
Me-based FI	0.138	0.181	0.078	0.137	0.061
Or-based FI	0.131	0.165	0.100	0.140	0.060

TABLE V. TESTING ERROR IN DIFFERENT FUZZY INTEGRAL

Data Cluster	1	2	3	4	5
WA	0.505	0.683	0.752	0.498	0.490
CI	0.209	0.325	0.483	0.194	0.048
Ma-based FI	0.187	0.234	0.064	0.150	0.212
Me-based FI	0.185	0.236	0.064	0.145	0.054
Or-based FI	0.185	0.247	0.103	0.150	0.056

TABLE VI. TRAINING TIME IN DIFFERENT FUZZY INTEGRAL

Data Cluster	1	2	3	4	5
WA	6.589s	4.955s	7.752s	6.734s	4.010s
CI	22.906s	10.644s	183.96s	64.723s	96.047s
Ma-based FI	1709s	1567s	34387s	3737s	4017s
Me-based FI	1712s	1572s	34450s	3727s	4014s
Or-based FI	23.809s	10.960s	190.26s	67.722s	99.701s

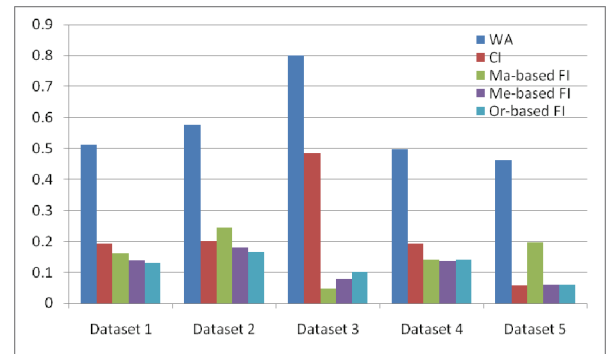


Figure 5. Training error in different Fuzzy Integral

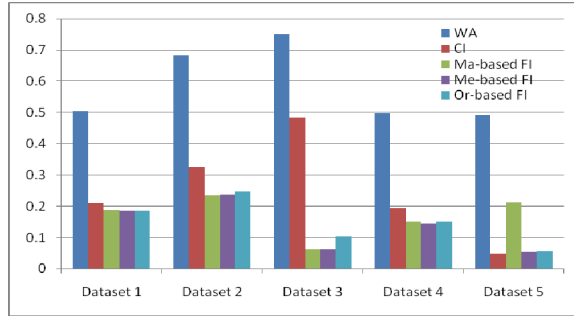


Figure 6. Testing error in different Fuzzy Integral

score from the game system. Training error, testing error and training time of our new fuzzy integrals compared with CI are summarized in TABLE IV-VI, figure 5, and figure 6.

From the experiment result, we provide the following conclusion:

- 1) Compared with weighted average, fuzzy integrals show an improvement in all the five datasets. Therefore, we believe that non-additive set function has a better representation than weighted average in predicting the power of unit combination.
- 2) Compared with Choquet Integral, all three new developed fuzzy integrals present better result in both training and testing. Me-based FI has a best performance with the average error of 0.128. Demonstrated by the result of dataset 3, the new fuzzy integrals are more effective if the dataset consists of various combinations. While Choquet Integral has a poor performance in this case.
- 3) The main weakness of Me-based FI and Ma-based FI is time consuming. Each time for set selection, they need to search for the whole fuzzy set. According table VI, we can observe that both of their training time is about 40 to 180 times more than Choquet Integral.
- 4) Jointly considering the training time and testing error, Or-based FI shows a best performance among these four methods. It has the similar training time with Choquet Integral with directly addressing the required set. Its training and testing error are similar with Me-based FI, and only 0.005 higher in average. We deem that Or-based FI is more appropriate in interaction detecting on Warcraft III data.

C. Usage Count of Set in Different Fuzzy Integrals

We tried to examine the difference of set selection in these four fuzzy integrals. All the selected fuzzy measures in the training process are being recorded and clustered with different amount of elements. As shown in table VIII, the selection of Or-based FI concentrates on the fuzzy measure with two to four elements. The count of selected set with more than five elements is dropped sharply. This phenomenon can be explained as: professional players seldom develop a balance army with many unit types. They usually focus on two to four unit types' cooperation. And this behavior make insufficient or

even no data contribute to train up the fuzzy measures with big number of elements.

Or-based FI selects the same amount of fuzzy measures with Choquet Integral, but different selecting result. As show in TABLE II, figure 7 and figure 8, we investigated that Or-based FI concentrates on fewer sets than Choquet Integral. Some fuzzy measures which represent the frequently used unit combination will have a higher chance to be selected by Or-based FI.

We analyze the set selection condition containing different number of elements. According to table VIII, the value of fuzzy sets with less than two elements is concentrating on two unit types which are necessary for each player to create primary. Then the fuzzy measure value concentrates on $\{1,2,3\}$, $\{1,2,4\}$, $\{1,2,6\}$ with three elements and $\{1,2,3,6\}$, $\{1,2,6,7\}$ with four elements. Obviously, it shows several different unit production sequences with different advanced units selection. Moreover, because the selection is based on unit production sequence, the battle with longer time which creates more unit types will not be affected by the shorter battle with less unit types. This also avoids the effect of unsuccessful strategy like player being defeated in the early stage and fails to produce the advanced units.

The table also shows that the advanced unit cannot produce alone in a battle. This is because every advanced unit requires additional resource and technology support. In another word, creating the advanced unit requires the contribution of basic unit. The advanced unit should dominate the army and cooperate with other units.

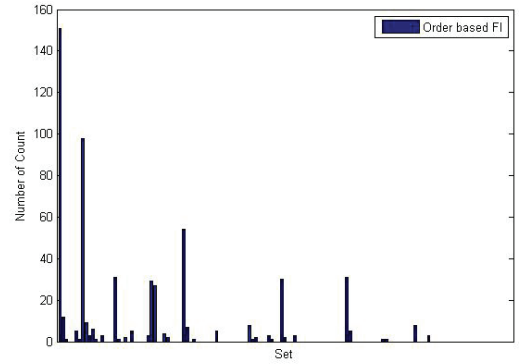


Figure 7. Count of Set selection for Order-based FI in Data cluster 1. X-axis is the set. Statring from the left, they are the set which only contains one element, i.e., $\{x_1\}$, $\{x_2\}$, $\{x_3\}$...etc. Then they are the sets which contain more elements, i.e., $\{x_1, x_2\}$, $\{x_1, x_3\}$, ..., $\{x_1, x_2, x_3\}$, etc. The last one on the right is the set which involves the whole set of elements. i.e., $\{x_1, x_2, \dots, x_n\}$

TABLE VII. COUNT OF THE FUZZY MEASURE SELECTION IN DATASET I

No. of variables in Set	Or-based FI	Choquet Integral
1	23.28571	23.85714
2	7.714286	8.000000
3	4.000000	3.942867
4	2.142857	2.085714
5	0.619408	0.666667
6	0	0

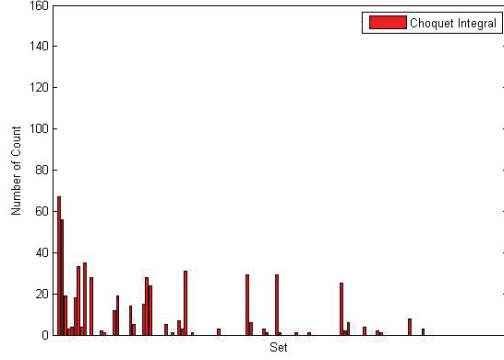


Figure 8. Count of Set selection for Choquet Integral in Data cluster 1

TABLE VIII. FUZZY MEASURE SELECTION IN DATASET I

Fuzzy Measure	Order based FI		Choquet Integral	
	Weighting	Number of count	Weighting	Number of count
{1}	0.460938	151	0.398438	0
{2}	0.617188	12	0.460938	67
{3}	0.734375	1	0.757813	56
{4}	0.859375	0	0.84375	19
{5}	0.789063	0	0.851563	3
{6}	0.296875	5	0.96875	4
{7}	0.578125	1	0.921875	18
{1,2}	0.453125	98	0.515625	33
{1,3}	0.453125	9	0.382813	4
{2,3}	0.601563	3	0.8125	35
{1,4}	0.789063	6	0.140625	0
{2,4}	0.359375	1	0.984375	28
{3,4}	0.429688	0	0.625	0
{1,5}	0.140625	3	0.664063	0
{2,5}	0.523438	0	0.929688	2
{3,5}	0.398438	0	0.046875	1
{4,5}	0.382813	0	0.734375	0
{1,6}	0.875	31	0.679688	0
{2,6}	0.914063	1	0.75	12
{3,6}	0.34375	0	0.976563	19
{4,6}	0.054688	2	0.1875	0
{5,6}	0.203125	0	0.34375	0
{1,7}	0.992188	5	0.242188	0
{2,7}	0.90625	0	0.976563	14
{3,7}	0.179688	0	0.875	5
...				
{6,7}	0.6875	3	0.96875	15
{1,2,3}	0.367188	29	0.960938	28
{1,2,4}	0.492188	27	0.992188	24
...				
{1,2,5}	0.835938	4	0.007813	0
{1,3,5}	0.34375	2	0.539063	0
{2,3,5}	0.625	0	0.835938	5
{1,4,5}	0.882813	0	0.359375	0
{2,4,5}	0.6875	0	0.875	1
{3,4,5}	0.6875	0	0.125	0
{1,2,6}	0.585938	54	0.695313	7
{1,3,6}	0.359375	7	0.789063	3
{2,3,6}	0.382813	0	0.992188	31
{1,4,6}	0.140625	1	0.8125	0
{2,4,6}	0.71875	0	0.375	1
...				
{1,2,7}	0.625	5	0.617188	0
{1,3,7}	0.835938	0	0.828125	0
{2,3,7}	0.632813	0	0.460938	3
...				

{1,6,7}	0.429688	8	0.90625	0
{2,6,7}	0.882813	1	0.875	29
{3,6,7}	0.554688	2	0.8125	6
...				
{1,2,3,5}	0.375	3	0.882813	3
{1,2,4,5}	0.421875	1	0.320313	1
...				
{1,2,3,6}	0.648438	30	0.992188	29
{1,2,4,6}	0.484375	2	0.859375	1
...				
{1,2,5,6}	0.007813	3	0.1875	0
{1,3,5,6}	0.585938	0	0.664063	0
{2,3,5,6}	0.0625	0	0.28125	1
...				
{1,2,3,7}	0.773438	0	0.25	1
...				
{1,2,6,7}	0.476563	31	0.992188	25
{1,3,6,7}	0.71875	5	0.882813	2
{2,3,6,7}	0.164063	0	0.851563	6
...				
{2,5,6,7}	0.679688	0	0.914063	4
...				
{1,2,3,4,6}	0.0625	1	0.140625	2
{1,2,3,5,6}	0.101563	1	0.625	1
...				
{1,2,3,6,7}	0.335938	8	0.945313	8
...				
{1,2,5,6,7}	0.546875	3	0.421875	3
...				
{1,2,3,4,5,6,7}	0.90625	0	0.460938	0

1 – Acolyte (Labor), 2 – Ghoul (Basic Melee unit), 3 – Crypt Fiend (Basic ranged unit), 4 – Gargoyle (Basic Air Unit), 5 – Meat Wagon (Advanced Siege Unit), 6 – Obsidian Statue (Advanced Support Unit), 7 – Destroyer (Advanced Air Unit). Number of count of Order based FI and CI = 0 are omitted in this table. Conclusion

V. CONCLUSION

In this paper, we apply different fuzzy integrals in predicting the power of unit combination in Warcraft III data. We found that weighted average cannot achieve an acceptable predicting accuracy. Since feature interaction exists in game data, non-additive method achieves better result. Furthermore, three new fuzzy integrals are developed with different directions to explain the game play. All of them attain a better result than the traditional Choquet Integral. Through the analysis of set selection, we have the evidence of why our methods got a better performance. We deem that for our data, Or-based FI has a better explanation of feature interaction in unit combination by considering the unit production sequence. Our future work regarding this topic includes new fuzzy application in micro-management. We attempt to generate corresponding unit movement to perform the selected strategy and finally lead the AI bots achieve the winning.

ACKNOWLEDGMENT

This research project is supported by the HK Polytechnic University grants 1-ZV5T, A-PJ18 and G-U523, and the NSFC grant no. 60903088.

REFERENCES

- [1] M. Sugeno, "Theory of fuzzy integrals and its applications," Ph.D. thesis, Tokyo Institute of Technology, 1974.
- [2] H. Tahani and H. J. M. Keller, "Information Fusion in Computer Vision using the Fuzzy Integral," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, pp. 733-741, 1990.

- [3] K. Hirota and W. Pedrycz, "Fuzzy computing for data mining," *Proceedings of the IEEE*, vol. 87, no. 9, pp. 1575 - 1600, 1999.
- [4] T. Murofushi, M. Sugeno and M. Machida, "Non-monotonic fuzzy measures and the Choquet integral," *Fuzzy Sets and Systems*, vol. 64, no. 1, pp. 73-86, 1994.
- [5] T. Murofushi, M. Sugeno, "The Choquet integral in multiattribute decision making," *Fuzzy measures and integrals: theory and applications*, 333-47, 2000.
- [6] D. Aha, M. Molineaux and M. Ponsen, "Learning to Win: Case-Based Plan Selection in a Real-Time Strategy Game," *Case-Based Reasoning Research and Development*, vol. 3620/2005, 2005.
- [7] J. L. Hsieh, and C. T. Sun, "Building a player strategy model by analyzing replays of real-time strategy games," *Proceedings of IEEE International Joint Conference on Neural Networks*, Hongkong, China, June 2008, pp. 3106-3111.
- [8] R. Miller, "Beyond ANOVA: basics of applied statistics, Chapman & Hall/CRC," 1997.
- [9] Z. Wang, K.S. Leung and J. Wang, "Genetic algorithms used for determining nonadditive set functions in information fusion," *Fuzzy Sets and Systems*, vol. 1, no. 9, pp. 518-521, 1997.
- [10] Z. Y. Wang, K. Xu and J. Wang, "Using genetic algorithm to determine non-negative monotone set functions for information fusion in environments with random perturbation," *International Journal of Intelligent System*, pp. 949-962, 1999.
- [11] Y.J. Li, Peter H. F. Ng, H. B. Wang, Y. Li and Simon C. K. Shiu, "Applying Fuzzy Integral for Performance Evaluation in Real Time Strategy Game," *Proceedings of 2010 2nd International Conference on Information and Multimedia Technology*, Hong Kong, December 28-30, 2010, pp. 168-172.
- [12] Peter H. F. Ng, Y. J. Li, H. B. Wang, Y. Li and Simon C. K. Shiu, "Bottom-Up Strategy Planning Model by applying Fuzzy Integral in RTS Game," *Proceedings of Joint 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems*, Okayama, Japan, 8-12 December 2010, pp. 8-12.
- [13] J. Wang and Z. Wang, "Using neural networks to determine Sugeno measures by statistics," *Neural Network*, pp. 183-195, 1997.
- [14] D. Yeung., X. Wang and E. Tsang, "Handling interaction in fuzzy production rule reasoning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, pp. 1979-1987, 2004.
- [15] Peter H. F. Ng, Simon C. K. Shiu and H. B. Wang, "Learning Player Behaviors in Real Time Strategy Games from Real Data," *Proceedings of Twelve Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp 321-327, 2009.
- [16] Y. Dodge, *The Oxford Dictionary of Statistical Terms*, Oxford University Press. ISBN 0-19-920613-9, 2003.