Modeling the Dynamic Individual Activity Decision Behavior Through the Influence Diagram Model

Natachai Wongchavalidkul

The Cluster of Logistics and Rail Engineering, Faculty of Engineering, Mahidol University, Salaya, Phuttamonthon, Nakhon Pathom, Thailand
E-mail: natachai.won@mahidol.ac.th

***Abstract*—Using the Influence Diagram (IFD), this paper attempted to model the human decision behaviors on choosing their daily activities by considering both the features of human decisions and the expected utilities outcomes of decisions. The 2004 time use survey data from the National Statistical Office (NSO) which represents 2,573 individuals in Bangkok were utilized as the case study in the paper. As the results, there are 19 proposed IFD models based on types of populations and number of members in the household. Results from the model validation processes also presented that all 19 IFD models were successfully simulated the activities (ACT) and the decisions of whom in the household to do activities with (WHO). The IFD model is expected to be better represent individual activity decisions than the Classification and Regression Tree (CART) model and the Bayesian network model as its incorporated the utility scores in the decision processes.**

***Index Terms*—Influence Diagram Model, Classification and Regression Tree, Particle Swarm Optimization, Activity Decision Behavior.**

I. Introduction

Individual and family traveling behavior has been incorporated into transportation traveling demand model in several ways. Models are differently considered person traveling behavior based on either the modeling structure and framework or the data availability. In general, the traditional traveling demand model represents persons' traveling behavior in the part of the modal split and the traffic assignment. However, the model still represents the demand of persons in the aggregate level of geography, known as Traffic Analysis Zone (TAZ).

The new concepts of travel demand modeling are interested in more details of person decisions and their traveling behavior. Several types of models have been introduced among these models. The activity based travel demand model catches up both researchers and practitioners in the field. However, based on its details and complexities, further researches are needed in order to solve several limitations in its concepts.

The activity based travel demand model is one of the most complex travel demand models that are trying to analyze the travel demand based on the individual activities behavior. In brief, the model is disaggregating analyze the individual persons in the study area (rather than concentrating on the zonal behavior, TAZ). Hence, several models are required to apply in this concept from generating locations of individual in the study area, analyzing their activity patterns, to estimating their traveling route decisions. This paper concentrates on the problem of modeling individual activity patterns based on their dynamic decisions.

Human decisions are uncertainty and consequences. Decisions are made under circumstances and those with the highly familiar are usually selected. Beside, time is also the important factor affected to the decisions. Decisions could be made over times, at one single time step, or under the time constraint. Human are normally making decisions by considering these tree features: the uncertainty, familiarity and expertise, and time [2]. On the other hand, Russell and Norvig restated the written of French philosopher Arnauld, written in 1662, regarding to the logic of human decisions as follows [5].

"To judge what one must do to obtain a good or avoid an evil, it is necessary to consider not only the good and the evil in itself, but also the probability that it happens or does not happen; and to view geometrically the proportion that all these things have together."

The good and evil in the context can be considered as the utility. The utility is a single number that expresses the desirability of a state. In the other word, human preferences between choices can be captured by utilities. In general, each decision choices are uncertainty. Hence, the combination of utilities (preferences) and probabilities (uncertainty) of choice decisions is the expected utility. In summary, human are making their decision based on their familiarity and expertise, time, and the expected utility.

The prior discussions of human behavior are considered to be true for the daily activity decision behavior. Using the Influence Diagram (IFD), the purposed model in this paper attempted to follow the prior assumptions on human decision behaviors by considering both the features of human decisions (the uncertainty, familiarity and expertise, and time) and the expected utilities outcomes of decisions.

II. Theoretical Background

A. The Influence Diagram (IFD)

The influence diagram (IFD) is a Bayesian network with integrations of action nodes and utility nodes. It is a directed acyclic graph with three types of nodes including the decision node, the chance or probabilistic node, and the utility node. The decision node represents the decision variables (choices of actions). The chance node represents random variables. These variables are the same as those used in the Bayesian network. Finally, the utility node describes the utility value (function) of its parent node [5]. Additionally, in this study, the discrete IFD is proposed. A discrete influence diagram (N) = (X, G, P, U) consists of:

- a Direct Acyclic Graph (DAG) (G) with nodes (V) and directed link (E)
- a set of discrete random variables (XC) and discrete decision variables (XD) such that X is a set of XC and XD, represented by nodes of G
- a set of conditional probability distributions (P) or conditional distribution table (CPT) containing one distribution $P(X_v | X_{pa(v)})$, for each discrete random variables Xv
- A set of utility function, U, containing one utility function, u(Xpa(w)) for each node w in the subset $V_u \subset V$ of utility nodes [6]

Given the structure of the IFD as prior discussions, the set of expected utility (EU) can be computed from the diagram as the following equation, (1).

$$EU(X) = \prod_{X_v \in X_c} P(X_v | X_{pa(v)}) \sum_{w \in V_u} u(X_{pa(w)}) \quad (1)$$

In general, the alternative (a*) which has the highest expected utility is proposed as the decision alternative. It is also known as the maximum expected utility principal, as stated in equation (2).

$$a^* = \arg\max EU(X) \quad (2)$$

However, in order to apply results of the expected utility to the individual decision behavior, the maximum expected utility principal is relaxed. Instead of selecting the activity which has the maximum expected utility, the model randomly selects the activity which probability is based on the value of expected utility, as shown in equation (3). Hence, the model will be more flexible on the decisions rather than only restricting decisions to the maximum expected utility.

$$a^* \sim EU(X) \quad (3)$$

B. Particle Swarm Optimization (PSO)

Briefly, PSO is a population based stochastic optimization technique. Two specific inspiration of the algorithm are fish schools and bird flocks. These natural behaviors represent the collections of particles (elements) that synchronously move together. The particles are moving together, diverging from another, and then regrouping in order to find the best solution. The algorithm works in two main steps, calculating the particle velocity and updating particle position. In the other words, the particles are recalculated its velocities and decide whether to move to the new position or not. Also, the moving decision will be made only the new position is gained the particle's values. Applications of PSO algorithms can be found in Pulido and Coello [3] and El-Zeonkolv [1]. The main properties of the PSO algorithms using in this study can be summarized as follows.

- The position of each particle is the utility values of a given ACT and WHO, u(ACT, WHO). The u(ACT, WHO) positions were also constrained as its coordinate values must be between 0 and 1 and the sum of the coordination values in each particle must be equal to 1.
- The score function of each particle is the Chi-square value gathering from the Chi-square test between the set of generated activities from the model at the given particle's position, u(ACT, WHO), and the target activities from the training data. Further, there are two scoring values to track. First, the P_{best} value is the best score that each particle can be gathered. Second, the G_{best} value is the score of the particle that have the best (minimum) score comparing to other particles in the swarm.

The velocity of each particle is calculated based on the particle's current position and the position of the G_{best} particle. Equation (4) presents the velocity calculation using in this research.

$$\begin{aligned} V_{k+1} = w \times V_k + r_1 \times (POS\_P_{best,k} - x_k) + \\ r_2 \times (POS\_G_{best,k} - x_k) \end{aligned} \quad (4)$$

Where:

- ○ V_k is the current velocity of the particle, iteration k
- ○ r_1 and r_2 is random numbers between 0.1 and 0.2 [3]
- ○ w is the inertia weight, which this research took a value randomly generated within the range [0.1,0.9] [3]
- ○ POS_P_{best} is the particle's position which has the best score (P_{best})
- ○ POS_G_{best} is the position of the particle's which has the best score in the swarm (G_{best}).
- ○ x_k is the expected new particle position, $x_k = x_{k-1} + V_{k-1}$

The perturbing function is used to perturb the swarm at the beginning period of the iterations. The main objective is to avoid the problems of particles' trapping in the local optima. Recommended by Pulido and Coello [3], equation (5) presents the perturb function used in this research. Further, in this research the perturbed particle will randomly jump to the new position which is assigned by equation (6).

$$Prob_{turbulence} = (\frac{N_{iter}}{Total_{iter}})^{1.7} - 2.0 \times (\frac{N_{iter}}{Total_{iter}}) + 1.0 \quad (5)$$

Where:

- $Prob_{turbulence}$ is the probability of the perturbed action
- N_{iter} is the number of the current iteration
- $Total_{iter}$ is the total number of the iterations

$$X_{perturb} = 2 \times RandomNumber \times X_k \quad (6)$$

Where:

- $x_{perturb}$ is the new position of the perturbed particle.
- x_k is the expected new particle position, $X_k = X_{k-1} + V_{k-1}$

- The stopping criteria assigned in this study is the total number of iterations. The total number of iterations was assigned by trials and errors (1,000 iterations were assigned in this study) to ensure that the G_{best} value given by the swarm was steady, no sign of the improvement.

Further, the problem can also be formulated in a form of the nonlinear programming problem as follows.

$$U_{jk} = \arg\max \left[\sum_{k=1}^{2}\sum_{j=1}^{37}\left(\frac{\left(N_{jk}^{S} - E_{jk}^{S}\right)^2}{E_{jk}^{S}}\right) + \sum_{k=1}^{2}\sum_{j=1}^{37}\left(\frac{\left(N_{jk}^{T} - E_{jk}^{T}\right)^2}{E_{jk}^{T}}\right)\right] \quad (7)$$

Subject to:

$$0 < U_{jk} \leq 1U$$

$$\sum_{k=1}^{2}\sum_{j=1}^{37} U_{jk} = 1$$

Where:

j = variable indicate type of activities (1-37), ACT

k = variable indicate if the activity involve with persons in the household or not (1-2), WHO

U_{jk} is the utility values of the given variables, j and k

$E_{jk}^{S} = \left(\frac{N_{jk} \times S}{N}\right)$ is the expected frequency of the simulated population in category j and k

$E_{jk}^{T} = \left(\frac{N_{jk} \times T}{N}\right)$ is the expected frequency of the target population in category j and k

$N_{jk} = N_{jk}^{S} + N_{jk}^{T}$ is the total number of simulated population and target population by category j and k

N_{jk}^{S} is the total number of simulated population by category j and k

N_{jk}^{T} is the total number of target population by category j and k

$S = \sum_{k=1}^{2}\sum_{j=1}^{37} N_{jk}^{S}$ is the total simulated population

$T = \sum_{k=1}^{2}\sum_{j=1}^{37} N_{jk}^{T}$ is the total target population

$N = S + T$ is the total simulated population and target population

III. Methodology

A. DATA FOR MODEL DEVELOPMENT

The 2004 time use survey data from the National Statistical Office (NSO) were utilized in this research. The data were sampled from different parts of Thailand, including Bangkok metropolitan area, central region, northern region, north eastern region, and southern region. In this research, Bangkok time use data was utilized. Daily activity schedule of 2,573 individuals from 1,013 households were recorded in the dataset. Additionally, 15 percent are single person households; 37 percent are two person households; 23 percent are three person households; and 25 percent are more than three person households. The data also includes both population socio-demographic attributes and daily activity attributes

B. THE PROPOSED IFD MODEL

Figure 1 presents the proposed IFD for the individual activity decision. The structure of the model is similar to the Bayesian network's structure, except that the IFD was introduced both decision nodes and a utility node into its structure.

Additionally, Table I - Table III describe definitions of each variable.

For this research, the nodes named as ACT and WHO were introduced as the decision node and the utility node was introduced into the model. The structure was simple constructed to the concept that the Activity (ACT) and the involvement of persons in the household in those activities (WHO) are depended to both the socio-demographic variables (i.e. number of member in the household, relationship of the person in the household, gender, age, and educational level, etc.) and the activity schedule variables (i.e. times of the day, activity time durations, and types of activities, etc.). Also, the relation between ACT and WHO will affect to the individual utility during the day as well.

Furthermore, the node "TYPE" in this model is the Type of persons which were classified and grouped using Classification and Regression Tree Algorithms. The processes of grouping and classifying persons based on their activity patterns are out of the scope of this paper but readers who interest can see more details in N.Wongchavalidkul and M. Piantanakulchai [4].

Additionally, equation (7) presents the mathematical formula of the proposed IFD structure (Fig. I.). The equation is derived from the general formula presented in equation (1). Further, the proposed IFD is the discrete IFD which is only comprised of the discrete node type. Therefore, the continuous variables, such as TDU and AGE (Table II) are discretized.

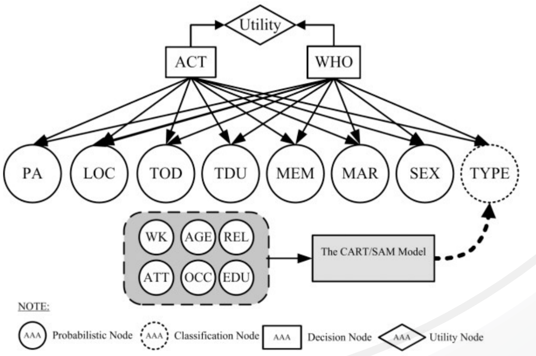

Fig. I. the Proposed IFD for Individual Activity Decision Behavior

$$EU(act, who \mid pa, loc, tod, tdu, mem, mar, sex, type) = P(pa \mid act, who)P(loc \mid act, who) P(tod \mid act, who)P(tdu \mid act, who)P(mem \mid act, who) P(mar \mid act, who)P(sex \mid act, who)P(type \mid act, who)u(act, who) \quad (8)$$

Given:
$act \in ACT$, $who \in WHO$, $pa \in PA$, $loc \in LOC$, $tod \in TOD$, $tdu \in TDU$, $mem \in MEM$, $mar \in MAR$, $sex \in SEX$, $type \in TYPE$
$X_D = \{ACT, WHO\}$ = *Decision Node*
$X_C = \{PA, LOC, TOD, TDU, MEM, MAR, SEX, TYPE\}$
= *Chance or Probabilistic Node*
$u(ACT, WHO)$ = *the utility of doing activity* (*ACT*) *with whom in the household* (*WHO*)

TABLE I
POPULATION SOCIO-DEMOGRAPHIC ATTRIBUTES

TABLE I
POPULATION SOCIO-DEMOGRAPHIC ATTRIBUTES

Variables	Description
PID	Personal ID (References for Survey Database)
MEM	Number of household members (age 10 years and over)
REL	Relationship of person in the household: 1. Householder, 2. Wife/Husband, 3. Single Son/Daughter, 4. Married Son/Daughter, 5. Son-in-law/Daughter-in-law, 6. Son/ Daughter's children, 7. Father/Mother/ Grandfather/Grandmother, 8. Relatives 9. Lodger or Servant
SEX	Gender: 1. Male, 2. Female
AGE	Age: 10-98 years old
MAR	Marital status: 1. Single 2. Married
ATT	School Attendance: 1. Attending School 2. Not Attending School
EDU	Education: 1. Illiterate, 2. Less than Lower Elementary, 3. Elementary Level, 4. Lower Secondary Level, 5. Upper Secondary Level (General Programs), 6. Upper Secondary Level (Vocational Programs), 7. Upper Secondary Level (Educational Programs), 8. Diploma Level (Academic Programs), 9. Diploma Level (Technical Vocational Programs), 10. Diploma Level (Vocational Programs), 11. University (Academic Programs), 12. University (Technical Vocational Program), 13. University (Educational Lines), 14. Others, 15. Unknown
WK	Working status: 1. Employed, 2. Unemployed, 3. Wait for Working Season (Mainly Agriculture), 4.Housekeeper, 5. Study, 6. Children/Elderly, 7. Illness or Disabled, 8. Vacation/Retirement, 9. Others, 10. Age less than 15 years old.
RG	Religion: 1. Buddhism, 2. Christianity, 3. Islam, 4. Others
OCC	Occupation: 1. Employers, 2. Business Owner with no Employees, 3. Working for Family Business (No Salary), 4. Government Employee, 5. State Enterprise Employee, 6. Private Firm Employee, 7. Group / Organization, 8. Others

TABLE II
ACTIVITY SCHEDULE ATTRIBUTES

Variables	Description
TOD	Time of the Day (1-24)
TDU	Activity Time Durations (1-144, 10 minutes per step)
WHO	The person is doing the activity with members in the household or not (1. No, 2.Yes)
ACT PA LOC	Activity Code: 1-37 (see table 5.3 for details) the previous activity: 1-37 (see table 5.3 for details) Activity locations: 1-9 (1.Home, 2. Home of Others, 3. Work Places, 4. Public Places, 5. Shop/Entertainment Venue, 6. Schools/Collages/Universities, 7. Cultural or Sport Places, 8.Vehicles, 9. Others)

TABLE III

ACTIVITY CODES (ACT) FOR MODEL DEVELOPMENT

Code	Description	Code	Description
1	Working in private firm, government, state enterprise, or organizations	21	Reading books
2	Lunch break	22	Reading Newspaper / Magazine
3	Traveling for activity type 1 and 2	23	Watching television
4	Working others (Business owners)	24	Listening to the radio
5	Traveling for activity type 4	25	Surfing internet
6	Cooking	26	Using library
7	Cleaning and Housekeeping	27	Traveling for activity type 19-26
8	Laundry	28	Sleep
9	Shopping	29	Taking nap
10	Taking care of household members and pets	30	Eating snack
11	Traveling for activity type 6-10	31	Social drinking
12	Conversation / Phone conversation	32	Self care
13	Community services and socialization	33	Chant
14	Traveling for activity type 13	34	Conversation and talk with family members (family times)
15	Study in school system	35	Do noting / Relax
16	Homework /review lectures / tutorial classes	36	Reflective thinking and planning
17	Self study / part-time courses	37	Traveling for activity type 28-36
18	Traveling for activity type 15-17		
19	Playing games (arcades, computers, etc.)		
20	Jogging / exercises		

Note: highlighted cells are the traveling activities

C. MODEL BUILDING AND CALIBRATION

Two main processes were taken for the influence diagram model developments, the model building and calibration process and the model validation process. The input data were divided into two dataset. Ninety percent of the survey data were randomly selected and used as the training or calibrating data. On the other hand, the other ten percent were later used as the testing or validating data. Fig. II presents the model development processes for the influence diagram model developed in this research.

The additional steps are the construction of the influence diagram structure and the calibration of the utility values. From equation (3), there are two types of parameters needed to be estimated in the model. The first parameter is the conditional probability distribution. The second parameter is the utility values. The conditional probability distribution (P), P(XC|XD), for each variable can be directly estimated from the training data using the message passing in a tree junction method.

However, the utility function, u(ACT, WHO), or the utility values of given ACT and WHO variables, cannot be interpreted directly from the training data. In this case, the Particle Swarm Optimization (PSO) method was applied to find the best utility values of the proposed influence diagram. The PSO method was selected over the other method because it is well known and the method has been successfully applied in many researches and applications. The method also proved that it gets results in a faster, cheaper way compared to other methods such as the genetic algorithm. Additionally, PSO is attractive because it only requires very few parameters to adjust. Hence, it is very simple to apply the method to study problems.

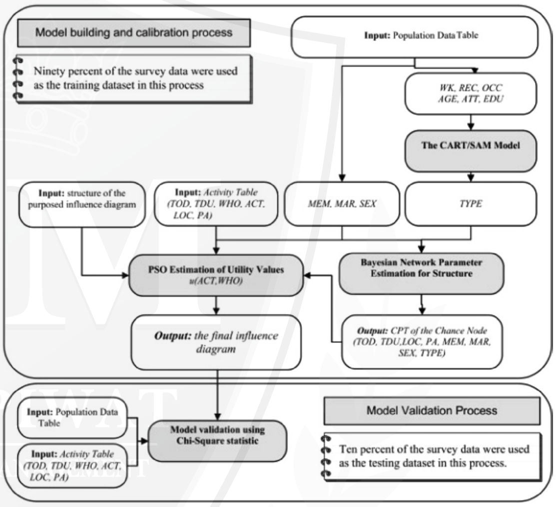

Fig. II. The IFD Model Development Processes

D. MODEL VALIDATION

The completed influence diagram developed in the previous section was then validated to qualify its ability in generating activity schedule. As mentioned in the previous section, the testing dataset was used in order to test performances of the model. In this research, the Chi-Square test was applied to compare the target activities in the testing dataset and the generated activities from the model. For the Chi-Square test, this research assumes that the P-value must be greater than 0.05 in order to accept the null hypothesis (Ho). The followings summarize the hypothesis testing used in the research.

Ho: *The distribution of the target populations and the simulated populations are not different*

$$\chi^2 = \sum_{k=1}^{2}\sum_{j=1}^{37}\left(\frac{\left(N_{jk}^{S}-E_{jk}^{S}\right)^2}{E_{jk}^{S}}\right)+\sum_{k=1}^{2}\sum_{j=1}^{37}\left(\frac{\left(N_{jk}^{T}-E_{jk}^{T}\right)^2}{E_{jk}^{T}}\right) \quad (9)$$

Note: Degree of freedom (df) = (r-1)(c-1) , df in this study = (72-1)(2-1) = 71

Where:

$E_{jk}^{S} = \left(\frac{N_{jk} \times S}{N} \right)$ is the expected frequency of the simulated population in category j and k

$E_{jk}^{T} = \left(\frac{N_{jk} \times T}{N} \right)$ is the expected frequency of the target population in category j and k

$N_{jk} = N_{jk}^{S} + N_{jk}^{T}$ is the total number of simulated population and target population by category j and k

N_{jk}^{S} is the total number of simulated population by category j and k

N_{jk}^{T} is the total number of target population by category *j* and *k*

$S = \sum_{k}^{2} \sum_{j}^{37} N_{jk}^{S}$ is the total simulated population

$T = \sum_{k}^{2} \sum_{j}^{37} N_{jk}^{T}$ is the total target population

$N = S + T$ is the total simulated population and target population

IV. Results of Model Development

From results of the calibration and validation process, it was found that the utility values calibrated from the proposed model structure in Fig. I. did not pass the validation process (P-value < 0.05). The reasons would be that the estimated utility values in the model were too general and could not be represented all types of persons in the model. In other words, the utility values are expected to be different from one type of persons to the others. Hence, the models were divided by types of population and individually calibrated for each type. Fig. III. presents the model structures for each population type. On the other hand, in order to pass the validation process, model for population type 7 was also divided by the variable indicated number of household members (MEM). Three categories of the variable MEM specified in the model of population type 7 are a single household (MEM = 1), a household with two members (MEM = 2), a household with three members (MEM = 3), and a household which has more than three members (MEM > 3).

Further, Fig. IV. presents the calibration results of the models for population type 1-9 (employment) and Fig. V. presents the calibration results of the models for population type 10-16 (unemployment). Table IV. summarized the final influence diagram developed in this research as well as the results from model validations.

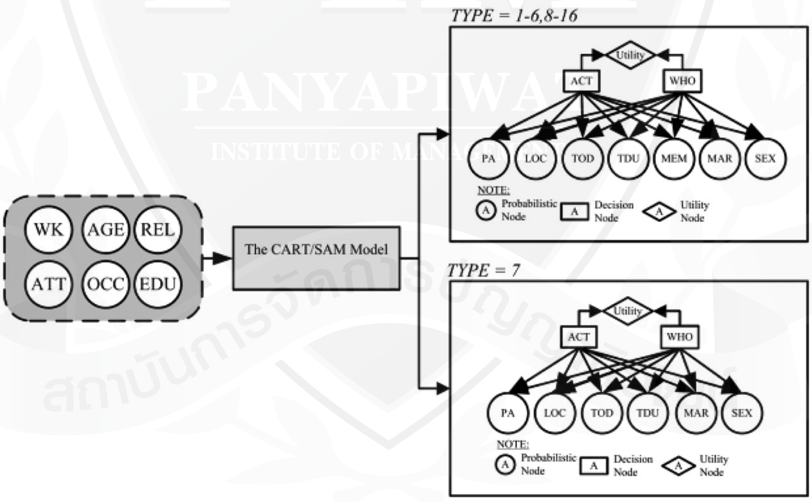

Fig. III. Final influence diagram structures

Note: for population type 7, the models were separately calibrated by variable MEM = 1, 2, 3, and more than 3

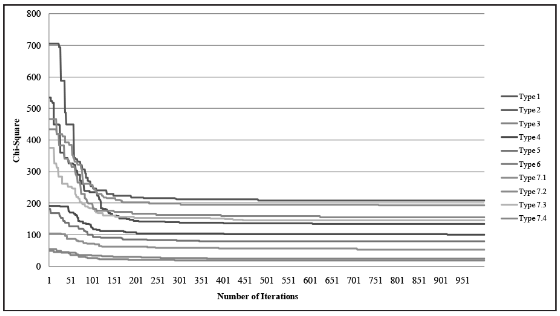

Fig. VI. Calibration Results of the Models for Population Type 1-9 (employed)

Note: type 7.1 (TYPE 7, MEM=1), type 7.2 (TYPE 7, MEM=2), type 7.3 (TYPE 7, MEM=3), type 7.4 (TYPE 7, MEM= more than 3)

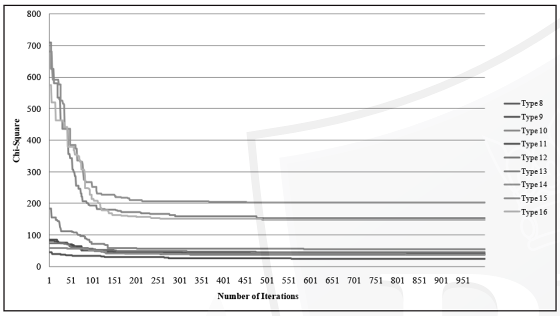

Fig. V. Calibration Results of the Models for Population Type 10-16 (unemployed)

TABLE IV

Summary of the Final Influence Diagrams and their Validation Results

Model	Variable		Chi-Square	P-Value
	TYPE	MEM		
1	1		61.103	0.782
2	2		50.284	0.966
3	3		14.874	1.000
4	4		37.758	0.999
5	5		29.631	1.000
6	6		15.892	1.000
7.1	7	1	18.175	1.000
7.2	7	2	52.169	0.950
7.3	7	3	41.570	0.997
7.4	7	> 3	50.732	0.966
8	8		21.099	1.000
9	9		16.670	1.000
10	10		52.348	0.939
11	11		30.208	1.000
12	12		25.954	1.000
13	13		32.076	1.000
14	14		43.798	0.996
15	15		28.730	1.000
16	16		48.860	0.962

Note: P-value must be greater than 0.05 in order to pass the validation processes

V. Summary and Discussions

The IFD model developed in this paper is targeted to model the decision behavior of the individual persons. As the results from the model development, it is found that different types of population require different sets of the utilities. Hence, there are 19 proposed IFD models based on types of populations and number of members in the household. Results from the model validation processes also presented that all 19 IFD models were successfully simulated the activities (ACT) and the decisions of whom in the household to do activities with (WHO).

The IFD model is expected to be better represent individual activity decisions than the Classification and Regression Tree (CART) model and the Bayesian network model as its incorporated the utility scores on the decision processes. However, practically, the performances of these models are not yet compared. In order to proceed to the conclusions, the comparisons of these modeling approaches should be provided, in the further research.

References

[1] A. M. El-Zonkoly , "Optimal placement of multi-distributed generation units including different load models using particle swarm optimization," Swarm and Evolutionary Computation, In Press, Corrected Proof. http://doi.org/10.1016/j.swevo.2011.02.003

[2] C. D. Wickens and Hollands, J. G., "Engineering Psychology and Human Performance (3rd Edition)," Prentice Hall Inc., 2000.

[3] G. T. Pulido and C. A. C. Coello "A constraint-handling mechanism for particle swarm optimization," In Evolutionary Computation, 2004, vol. 2, pp. 1396-1403.

[4] N. Wongchavalidkul and M. Piantanakulchai, "The Integration of Classification Tree and Sequence Alignment Method for Exploring Groups of Population Based on Daily Time Use Data," Applied Soft Computing, 2015, vol. 34, pp. 106-119.

[5] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach (2nd Edition)," Prentice Hall., 2003.

[6] U. B. Kjarulff and A. L. Madsen, "Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis," Springer., 2008.

Acknowledgement

A part of this research was funding by the SIIT's graduate scholarship. The author is also would like to thank you Assoc. Prof. Dr. Mongkut Piantanakulch for his advising all over the topic.

Natachai W. is currently a lecturer in Cluster of Logistics and Rail Engineering at Engineering Faculty, Mahidol University, Thailand. He received M.Sc. in Civil Engineering (Transportation) Howard R. Hughes College of Engineering. School of Civil and Environmental Engineering and Construction. University of Nevada Las Vegas. He also received his Ph.D. in Engineering (Transportation) from Sirindhorn International Institute of Technology, Thammasat University