A Kind of Risk-Sensitive Group Decision-Making Based on MDP

Mingshan You, Wei Zeng(corresponding author), Hongtao Zhou

Institute of Systems Engineering, Huazhong University of Science and Technology,

Wuhan 430074, P.R.China

[1]ymsyouyou@163.com,[2]zengwei@mail.hust.edu.cn,[3]zht730@yahoo.com.cn

Abstract. One-switch utility function is used to describe how the risk attitude of a decision maker changes with his wealth level. In this paper additive decision rule is used for the aggregation of decision member's utility which is represented by one-switch utility function. Based on Markov decision processes (MDP) and group utility, a dynamic, multi-stages and risk sensitive group decision model is proposed. The proposed model augments the state of MDP with wealth level, so the policy of the model is defined as an action executed in a state and a wealth level interval. A backward-induction algorithm is given to solve the optimal policy for the model. Numerical examples show that personal risk attitude has a great influence on group decision-making when personal risk attitudes of members are different, while the weights of members play a critical role when personal risk attitudes of members are similar.

Introduction

When facing major decisions concerning heavy casualties and property losses, decision maker's risk attitude is generally risk-averse. Exponential utility function is usually used to model risk-averse attitude. The risk attitude of the decision maker is modeled by exponential utility function keeps unchanged. However, in reality, personal risk attitude generally change with wealth level [1]. Bell gives a kind of utility function named one-switch utility function, that can describe the risk attitude of decision makers whose risk attitude shift from risk-averse to risk-neutral as their wealth level increases. Furthermore Bell and Fishburn give a kind of one-switch utility function that meets SOS(strong one-switch) conditions, that is formed by linear utility function and exponential utility function, $u(w) = aw - be^{-cw}$ (where w is wealth level). Liu and Koenig improve the one-switch utility function with considering the risk aversion coefficient γ, that is $u(w) = w - D\gamma^w$. This form of one-switch function can not only describe the change of decision maker's risk attitude, but also can distinguish different decision maker's risk attitude through the quantitative risk aversion coefficient γ [2-6].

For a multi-stages and uncertain decision problem, if the decision maker is risk-sensitive, it is necessary to consider the possible change of decision maker's risk attitude when the wealth level of decision maker changes and further influence on the decision in the next stage. Markov decision processes (MDP) is an important model framework for uncertain and multi-stages decision theoretic planning. Based on MDP, Howard and Matheson use exponential utility function to depict the risk attitude of decision makers and maximize the expected utility instead of maximizing the expected reward to seek the optimal policy, but they do not consider the wealth level of decision makers [7].

For the study of risk-sensitive group decision making, the general method is assembling the personal utility function into group utility function and making decisions based on the group utility function. The aggregation methods include addition rule and multiplication rule, etc [8-10]. Based on multi-goal or multi-attribute utility functions, single stage and static decision making is focused on in most studies without considering the multi-stages, dynamic group decision making.

In this paper, we will study a dynamic, multi-stages and uncertain group decision making in which all the members are risk sensitive. Group utility function is proposed based on one-switch utility function. MDP model of risk-sensitive group decision making is given and backward

induction method is used to solve the model. A numerical example is illustrated to analyze the impact of the personal risk attitude on group decision making.

Risk-Sensitive Group Decision Making Model Based On MDP

One-Switch Utility Function. One-switch utility function denotes that there existing a wealth level w, when the wealth level of a decision maker is below w, the decision maker is risk-averse, but when his wealth level is higher than w, the decision maker is approach to risk-neutral. The form of one-switch utility function given by Liu and Koenig is shown as follows [4]:

$$u(w) = w - D\gamma^w, \ D>0, \ 0<\gamma<1 \tag{1}$$

Where w is wealth level, γ is risk aversion coefficient of the decision maker. D is constant coefficient provides an adjustable tradeoff between linear term and exponential term. In this paper, facing heavy casualties and property losses, w is a large negative number, then D is a very small number close to 0. The one-switch function is illustrated in Fig. 1.
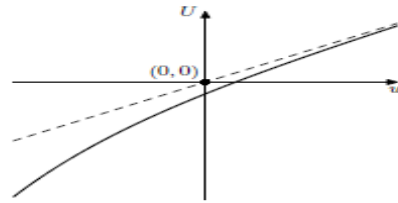

Fig. 1: One-switch utility function

If value function in MDP is replaced with one-switch utility function, the planning problem will be not decomposable [1]. Traditional value iteration method cannot be used to calculate the expected utility and seek the optimal policies. A new backward-induction method of risk-sensitive group decision making is given in the following sections.

Risk-Sensitive Group Decision Making Model Based on MDP. Goal-directed Markov decision processes (GDMDP) that MDP has goal states is named augmented GDMDP if wealth level is considered as a part of the state of the original GDMDP. So the state of augmented GDMDP is composed by the states of original GDMDP and wealth level. The state set is (S, W), goal state set is (G, W) and non-goal state set is (S', W), where S denotes state set original GDMDP, W denotes wealth level set and $S' = S \backslash G$.

Action set is A_s(the same as the action set of original GDMDP), which denotes the actions that decision maker can execute in the state s.

$s_t \in S, a_t$ and w_t denote the system state, action and wealth level in the stage t respectively. If system state is s_t, decision maker executes action a_t, and the system state is s' in the next stage, then the state transition probability and reward are represented respectively

$$P\big((s', w_{t+1})\big|(s_t, w_t), a_t\big) = \begin{cases} P(s'|s_t, a_t), & w_{t+1} = w_t + r(s_t, a_t, s') \\ 0, & other \end{cases}$$

$$r_t((s_t, w_t), a_t, (s', w_{t+1})) = u(w_{t+1}) - u(w_t) < 0 \tag{2}$$

In this paper, we use additive decision rule for the aggregation of utility $u(w) = \sum_{i=1}^{n} k_i u_i(w)$, where $\sum_{i=1}^{n} k_i = 1$. $u_i(w) \in [0,1]$ denotes the ith member's one-switch utility function. k_i is the weights of $u_i(w)$, γ_i is the risk aversion coefficient of ith member. $i = 1,2, \dots, n$, n is the number of group members.

$$u(w) = \sum_{i=1}^{n} k_i u_i(w) = w - k_1 D_1 \gamma_1{}^w - k_2 D_2 \gamma_2{}^w - \cdots - k_n D_n \gamma_n{}^w \tag{3}$$

In any stage, for a policy π the expected group utility in the state (s, w) is expressed as follows:

$$\begin{cases} v^\pi(s,w) = u(w) = w - k_1 D_1 \gamma_1{}^w - k_2 D_2 \gamma_2{}^w - \cdots - k_n D_n \gamma_n{}^w & \forall s \in G, \forall w \\ v^\pi(s,w) = \sum_{s' \in S} P(s'|s, \pi(s,w)) \cdot v^\pi(s', w + r(s, \pi(s,w), s')) & \forall s \in S', \forall w \end{cases} \tag{4}$$

So the optimal policy is expressed as follows:

$$\begin{cases} v^*(s,w) = u(w) = w - k_1 D_1 \gamma_1{}^w - k_2 D_2 \gamma_2{}^w - \cdots - k_n D_n \gamma_n{}^w & \forall s \in G, \forall w \\ v^*(s,w) = \max_{a \in A_s} \sum_{s' \in S} P(s'|s, a) \cdot v^*(s', w + r(s, a, s')) & \forall s \in S', \forall w \end{cases} \tag{5}$$

As wealth level being a part of state, the policy π is defined as the action executed in a state s and wealth level w: $\pi = ((s,w), a)$. Wealth level is a continuous interval $(-\infty, w_0]$ (w_0 denotes initial wealth level, and we set $w_0 = 0$), so it is necessary to divide the wealth level interval into child intervals. Then policy is the action executed in a state s and wealth level interval $(w^i, w^{i+1}]$, denoted as $\left((s, (w^i, w^{i+1}]), a\right)$ (w^i denote a wealth level point).

The Backward-Induction Algorithm

In group utility function, personal utility function is usually normalized, $u_i(w) \in [0,1]$. For the one-switch utility function, $u_i(w) = w - D_i \gamma_i{}^w$, if the value of w is in the range of from $-\infty$ to 0, the value of $u_i(w)$ will be in the range $(-\infty, D_i]$, because D_i is a very small number close to 0. The value ranges of all $u_i(w)$ approach to the same interval $(-\infty, 0)$, it is not necessary to normalize the utility function.

For a policy π the expected group utility in the state (s, w) is calculated as follows:

$$\begin{aligned} v^\pi(s,w) &= \lim_{t \to \infty} E^\pi_{s,w}[u(w_t)] \\ &= \lim_{t \to \infty} E^\pi_{s,w}[u_l(w_t) + k_1 D_1 u_{1,e}(w_t) + k_2 D_2 u_{2,e}(w_t) + \cdots + k_n D_n u_{n,e}(w_t)] \\ &= \lim_{t \to \infty} E^\pi_{s,w}[u_l(w_t)] + \lim_{t \to \infty} E^\pi_{s,w}[k_1 D_1 u_{1,e}(w_t)] + \cdots + \lim_{t \to \infty} E^\pi_{s,w}[k_1 D_1 u_{1,e}(w_t)] \\ &= \lim_{t \to \infty} E^\pi_{s,w}\left[w + \sum_{i=0}^{t-1} r_i\right] + \lim_{t \to \infty} E^\pi_{s,w}\left[k_1 D_1 \gamma_1{}^{w + \Sigma_{i=0}^{t-1} r_i}\right] + \cdots + \lim_{t \to \infty} E^\pi_{s,w}[k_n D_n \gamma_n{}^{w + \Sigma_{i=0}^{t-1} r_i}] \\ &= w + \lim_{t \to \infty} E^\pi_{s,w}\left[\sum_{i=0}^{t-1} r_i\right] + k_1 D_1 \gamma_1{}^w \lim_{t \to \infty} E^\pi_{s,w}\left[\gamma_1{}^{\Sigma_{i=0}^{t-1} r_i}\right] + \cdots + k_n D_n \gamma_n{}^w \lim_{t \to \infty} E^\pi_{s,w}\left[\gamma_n{}^{\Sigma_{i=0}^{t-1} r_i}\right] \\ &= w + v_l^\pi(s) + k_1 D_1 \gamma_1{}^w v_{1,e}^\pi(s) + \cdots + k_n D_n \gamma_n{}^w v_{n,e}^\pi(s) \\ \therefore v^\pi(s,w) &= w + v_l^\pi(s) + k_1 D_1 \gamma_1{}^w v_{1,e}^\pi(s) + \cdots + k_n D_n \gamma_n{}^w v_{n,e}^\pi(s) \end{aligned} \tag{6}$$

$v_l^\pi(s)$, $v_{i,e}^\pi(s)$ denote the linear expected utility and exponential expected utility respectively. From Eq.6, we can find the relationship between expected group utility and linear expected utility or exponential expected utility. For a state s and a policy π, linear expected utility and exponential expected utility can be calculated as follows:

$$\begin{cases} v_l^\pi(s) = 0 & s \in G \\ v_l^\pi(s) = \sum_{s' \in S} P(s'|s, \pi(s))[r(s, \pi(s), s') + v_l^\pi(s')] & s \in S' \end{cases} \tag{7}$$

$$\begin{cases} v_{i,e}^\pi(s) = -1 & s \in G \\ v_{i,e}^\pi(s) = \sum_{s' \in S} P(s'|s, \pi(s))[\gamma_i{}^{r(s, \pi(s), s')} \cdot v_{i,e}^\pi(s')] & s \in S' \end{cases} \tag{8}$$

Using Eq. 7 and Eq. 8, we can calculate the linear expected utility and exponential expected utility of all stationary policies when $t \to \infty$ or $w_t \to -\infty$. Then using Eq. 6, we can get the expected group utility. In details, value iteration algorithm is used to get the

values $v_l^\pi(s)$ and $v_{i,e}^\pi(s)$ through Eq.7 and Eq. 8, then the right hand side of Eq. 6 is a polynomial of one indeterminate about wealth level w. Through comparing the value of this polynomial under different policies, the optimal policy which has the biggest value of the polynomial can be obtained.

π^* is used to denote the optimal policy for one-switch utility function, $\pi_{i,e}^*$ denotes the optimal policy of ith member for exponential utility function. According to the backward induction idea of value iteration, we first seek the optimal policy $\pi_{-\infty}^*$ when $w \to -\infty$.

LEMMA 1. If there exist a γ_i, for all $j \neq i$, $\gamma_i < \gamma_j$, that is to say, the ith member is more risk-averse than other members, we have

$$\lim_{w \to -\infty} v^*(s,w)\gamma_i^{-w} = k_i D_i v_{i,e}^*(s) \tag{9}$$

Proof: $lim_{w\to-\infty} v^*(s,w)\gamma_i^{-w} = lim_{w\to-\infty} v^{\pi^*}(s,w)\gamma_i^{-w}$

$$= \lim_{w \to -\infty} w\gamma_i^{-w} + v_l^{\pi^*}(s)\gamma_i^{-w} + k_1 D_1 \left(\frac{\gamma_1}{\gamma_i}\right)^w v_{1,e}^{\pi^*}(s) + \cdots + k_i D_i v_{i,e}^{\pi^*}(s) +$$

$$\cdots + k_n D_n \left(\frac{\gamma_n}{\gamma_i}\right)^w v_{n,e}^{\pi^*}(s)$$

$\because \gamma_j/\gamma_i > 1, j \neq i$, so $\lim_{w\to-\infty} k_j D_j \left(\frac{\gamma_j}{\gamma_i}\right)^w v_{j,e}^{\pi^*}(s) = 0$

$\therefore \lim_{w\to-\infty} v^*(s,w)\gamma_i^{-w} = k_i D_i v_{i,e}^{\pi^*}(s) \leq k_i D_i v_{i,e}^*(s)$

and $\lim_{w\to-\infty} v^*(s,w)\gamma_i^{-w} \geq \lim_{w\to-\infty} v^{\pi_{i,e}^*}(s,w)\gamma_i^{-w}$

$$= \lim_{w \to -\infty} w\gamma_i^{-w} + v_l^{\pi_{i,e}^*}(s)\gamma_i^{-w} + k_1 D_1 \left(\frac{\gamma_1}{\gamma_i}\right)^w v_{1,e}^{\pi_{i,e}^*}(s) + \cdots + k_i D_i v_{i,e}^{\pi_{i,e}^*}(s) +$$

$$\cdots + k_n D_n \left(\frac{\gamma_n}{\gamma_i}\right)^w v_{n,e}^{\pi_{i,e}^*}(s)$$

$$= k_i D_i v_{i,e}^*(s)$$

$\therefore \lim_{w\to-\infty} v^*(s,w)\gamma_i^{-w} = k_i D_i v_{i,e}^*(s)$ ∎

Lemma 1 shows that, when the wealth level approaches to $-\infty$, the optimal policy is the same as the most risk-averse member's optimal policy $\pi_{i,e}^*$. The optimal policy $\pi_{i,e}^*$ is solved by calculating of MEUe (maximizing expected exponential utility).

After getting $\pi_{-\infty}^*$, the next step is finding out the wealth level threshold $\underline{w}$ which means that the optimal policy $\pi_{-\infty}^*$ keeps in the interval $(-\infty, \underline{w}]$. According to Eq. 5, for all $w \leq \underline{w}$, and all non-goal state s

$\because w + r(s,a,s') < \underline{w}$,

$v^*(s,w) = max_{a \in A_s} \sum_{s' \in S} P(s'|s,a) \cdot v^{\pi_{-\infty}^*}(s', w + r(s,a,s'))$

$$= max_{a \in A_s} \sum_{s' \in S} P(s'|s,a) \cdot (w + r(s,a,s') + v_l^{\pi_{-\infty}^*}(s) + k_1 D_1 \gamma_1^{w+r(s,a,s')} v_{1,e}^{\pi_{-\infty}^*}(s) + \cdots$$

$$+ k_n D_n \gamma_n^{w+r(s,a,s')} v_{n,e}^{\pi_{-\infty}^*}(s)$$

$$= max_{a \in A_s}[w + v_l^{\pi_{-\infty}^*}(s,a) + k_1 D_1 \gamma_1^w v_{1,e}^{\pi_{-\infty}^*}(s,a) + \cdots + k_n D_n \gamma_n^w v_{n,e}^{\pi_{-\infty}^*}(s,a)]$$

$v_l^{\pi_{-\infty}^*}(s,a)$ and $v_{i,e}^{\pi_{-\infty}^*}(s,a)$ can be calculated through Eq. 7and Eq. 8. $\forall s \in S'$, $\forall a_j \in A_s \backslash \pi_{-\infty}^*(s)$:

$\sum_{s' \in S} P(s'|s, \pi_{-\infty}^*(s)) \cdot v^{\pi_{-\infty}^*}(s', w + r(s, \pi_{-\infty}^*(s), s')) \geq$

$$\sum_{s' \in S} P(s'|s, a_j) \cdot v^{\pi_{-\infty}^*}(s', w + r(s, a_j, s'))$$

$w + v_l^{\pi_{-\infty}^*}(s, \pi_{-\infty}^*(s)) + k_1 D_1 \gamma_1^w v_{1,e}^{\pi_{-\infty}^*}(s, \pi_{-\infty}^*(s)) + \cdots + k_n D_n \gamma_n^w v_{n,e}^{\pi_{-\infty}^*}(s, \pi_{-\infty}^*(s)) \geq$

$$w + v_l^{\pi_{-\infty}^*}(s, a_j) + k_1 D_1 \gamma_1^w v_{1,e}^{\pi_{-\infty}^*}(s, a_j) + \cdots + k_n D_n \gamma_n^w v_{n,e}^{\pi_{-\infty}^*}(s, a_j) \tag{10}$$

We can get a wealth level point $\underline{w}_{\pi^*_{-\infty},s,a_j}$ in equality case of the weak inequality above.

From the algorithm above, we can get the wealth level threshold $\underline{w}$

$$\underline{w} = min_{s \in s'} \, min_{a_j \in A_s \backslash \pi^*_{-\infty}(s)} \, min \, \left(w_0, \underline{w}_{\pi^*_{-\infty},s,a_j}\right) \tag{11}$$

After getting $\underline{w}$, the next step is to divide further the interval $(\underline{w}, w_0]$ into child intervals and seek the optimal policy for each child interval. The main procedures of backward-induction algorithm for group decision making is listed as follows.

Step 1: For all $s \in s'$, get $\pi^*_{-\infty}$ and $v^*_{1,e}, v^*_{1,e} \cdots, v^*_{1,e}, v^*_l$ when $w \to -\infty$.

Step 2: For all $s \in s', a \in A_s \backslash \pi^*_{-\infty}(s)$, using Eq. 7 and Eq. 8 to get the value $v_l^{\pi^*_{-\infty}}(s,a)$, $v_{1,e}^{\pi^*_{-\infty}}(s,a), \cdots v_{n,e}^{\pi^*_{-\infty}}(s,a)$, then get $\underline{w}_{\pi^*_{-\infty},s,a}$.

Step 3: $\underline{w} = min_{s \in s'} \, min_{a_j \in A_s \backslash \pi^*_{-\infty}(s)} \, min \, \left(w_0, \underline{w}_{\pi^*_{-\infty},s,a_j}\right)$.

Step 4: For all $s \in s'$, dividing the interval $(\underline{w}, w_0]$. Inserting two different kinds of wealth level points and seek the optimal actions [4].

Numerical Example

For a serious forest fire, there are two ways to extinguish the fire, one is direct extinguishing, and the other one is digging a fire isolation zone. These two ways consume r_1 and r_2 units of resources with success possibility p_1 and p_2 respectively. Now there are two decision members named member1 and member2 with risk aversion coefficient γ_1 and γ_2 respectively. They need to make an optimal firefighting policy to extinguish the fire together.

For the above example, GDMDP model is defined as follows: S={on fire (S_0), fire extinguished (g)}, A_s={direct extinguishing (a_1), digging an fire isolation zone (a_2)}. Assuming state transition possibility of success $P = \{0.95, 1\}$, returns $R = \{-1000, -5000\}$, the one-switch utility functions of two members are $u_1(w) = w - 10^{-9}0.997^w$ and $u_2(w) = w - 10^{-9}0.995^w$. Then the optimal policies and wealth points of two members are represented as follows:

$$\text{Member1}: \pi^*(s_0, w) = \begin{cases} a_1 & w \in (-6211, 0] \\ a_2 & w \in (-\infty, -6211] \end{cases}$$

$$\text{Member2}: \pi^*(s_0, w) = \begin{cases} a_1 & w \in (-4077.2, 0] \\ a_2 & w \in (-\infty, -4077.2] \end{cases}$$

If set the weights of utility function of two members $k_1 = k_2 = 0.5$, then the optimal policies and wealth points are listed as follows:

$$\pi^*(s_0, w) = \begin{cases} a_1 & w \in (-4215.2, 0] \\ a_2 & w \in (-\infty, -4215.2] \end{cases}$$

The result shows that in the interval $(-6211, -4215.2]$, if member1 make decisions alone, the optimal policy is taking action a_1 in state S_0, and a_2 for member2, but if they make decisions together, then action a_2 is taken.

Next, we consider how the group-wealth level threshold $\underline{w}$ changes with the weights of two utility functions, in detailed, k_1 changes from 0.05 to 0.95 and k_2 changes from 0.95 to 0.05. The result is shown in Fig.2.

The result shows that $\underline{w}$ tends to the wealth point of member2 as a whole, that is, if the risk aversion coefficients of members are different, the influence of the weights is not very various, the risk aversion coefficients play an important role.

If the risk aversion coefficients of two members are set 0.99694 and 0.9969 respectively, the corresponding one-switch utility functions are $u_1(w) = w - 10^{-9}0.99694^w$ and $u_2(w) = w - 10^{-9}0.9969^w$. The weights of utility functions change as the same above, the result is shown as Fig.3. The result shows that, if there is little difference between the risk aversion coefficients of members, the weights of members play a critical role.
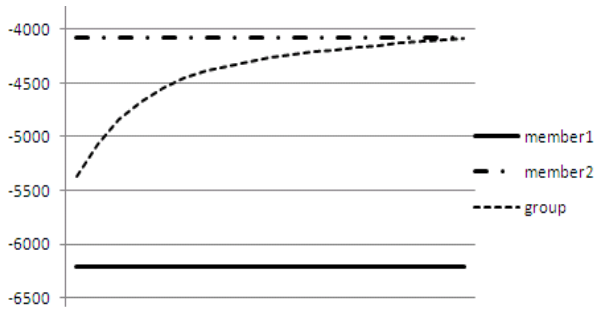
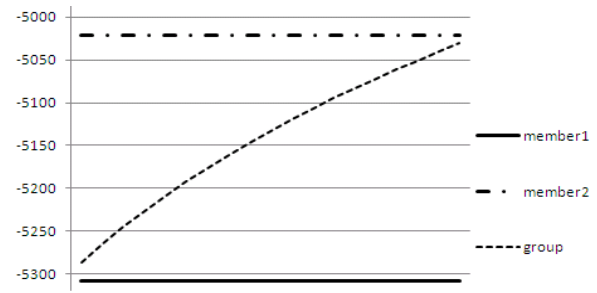| Fig. 2: The change of $\underline{w}$ (different risk aversion coefficients of members) | Fig. 3: The change of $\underline{w}$ (similar risk aversion coefficients of members) |

Conclusions

In a dynamic multi-stage group decision making, the risk attitudes of members are taken into consideration. A group decision model based on MDP and one-switch utility function and its algorithm are proposed. The influence of risk aversion coefficients of members is not various if the risk aversion coefficients have little difference. However, the risk aversion coefficients of members play a critical role if the risk aversion coefficients of members are different, the member with smaller risk aversion coefficient has larger influence on group decision making, and the weights of members have less impact on group decision making.

Acknowledgements

References

[1] Y. Liu, S. Koenig, Risk-sensitive planning with one-switch utility functions: Value iteration, in Proceedings of the Twentieth National Conference on Artificial Intelligence (2005) 993-999.

[2] D. E. Bell, One-switch utility functions and a measure of risk, J. Management Science. 34(1988) 1416-1424.

[3] D. E. Bell, P.C. Fishburn, Strong one-switch utility, J. Management Science. 47(2001) 601 -604.

[4] Y. Liu, S. Koenig, An exact algorithm for solving MDPs under risk-sensitive planning objectives with one-switch utility functions, in Proc. of 7th Int. Conf. on Autonomous Agents and Multi-agent Systems(2008).

[5] R.L. Winkler, Multiattribute one-switch utility, J. Management Science.   58(2012) 602-605.

[6] A.E. Abbas, D.E. Bell, One-Switch Independence for Multiattribute Utility Functions, J. Operations Research. 59(2011) 764–771.

[7] R.A. Howard, J.E. Matheson, Risk-sensitive Markov decision process, J. Management Science. 18(1972) 356-369.

[8] D.C. Morais, A.T.de Almeida, Group decision making on water resources based on analysis of individual rankings, J. Omega. 40(2012) 42-52.

[9] J. Xu, Z. Wu, A discrete consensus support model for multiple attribute group decision making, J. Knowledge-Based Systems. 24(2011) 1196-1202.

[10] M. Dursun, E.E. Karsak, M.A. Karadayi, A fuzzy multi-criteria group decision making framework for evaluating health-care waste disposal alternatives, J. Expert Systems with Applications. 38(2011) 11453-11462.