

The Internet Public Opinion Propagation Model in Uncertain Environment

Xin Gao^{1, a}, Lin Fu^{1, b} and Dan A. Ralescu^{2, c}

¹School of Mathematical Sciences and Physics, North China Electric Power University, Beijing 102206, China;

² Department of Mathematical Sciences, University of Cincinnati, OH 45221-0025, USA.

^agaoxin@ncepu.edu.cn, ^bncepufl@126.com, ^cdan.ralescu@math.uc.edu

Keywords: Internet public opinion, Uncertain programming, Uncertainty theory, Genetic algorithm.

Abstract. With the rapid development of the internet, the classical internet public opinion (IPO) problem as an important social issue has been studied for several years. However, few of models are out of the dimension of differential equations. In this paper, a novel mathematical model for IPO problem based on uncertainty theory is first proposed. A hybrid intelligent algorithm consisted by 99-method and an improved genetic algorithm is given to solve the proposed model. Finally, a numerical example based on a real event is given to show the efficiency and usefulness of the proposed methodology. The result shows that the IPO model achieves good modeling and control performance.

1. Introduction

Nowadays, the internet public opinion (IPO) problem becomes a hot issue studied by many scholars. According to China Internet Network Information Center (CNNIC) 39th statistical report on internet development in Beijing (released on the 22nd of January, 2017) it follows that-as of the end of December, 2016, Chinese netizens reached 731 million, the internet penetration rate reached 53.2%, thus about half of the Chinese population have access to the internet. Online social networks create a platform for people to communicate and connect, meanwhile, it has some characteristics by the development of the internet such as openness, complexity, and non integrity.

Under the relatively weak management of the network environment, the harmfulness of internet public opinion is magnified out of control. In order to solve this problem, Sudbury [1] first built SIR model based on the biological theory. After that, many scholars have investigated under different complex networks [2-9]. Meanwhile, Wang [10] applied game theory to the IPO problem model, Liu et al. [11] investigated the IPO problem by cellular automata.

Many of the above methods are using probability theory which are based on the existing data to deal with the uncertain factors, the IPO problem, however, exhibits sudden outbreaks with no historical data. In order to deal with such situations, uncertainty theory was introduced by Liu [12] and refined by Liu [13]. Uncertain differential equations [14,15], uncertain processes [16], uncertain programming [17,18], uncertain set theory [19,20], among others [21-24], have also been proposed. Later, Su [25] described the IPO problem under uncertain differential equation and combined it with real data examples to illustrate the effectiveness of the model.

This paper will propose a new mathematical model based on the previous studies and on uncertainty theory. Our model will extend the differential equations model mentioned above. A hybrid intelligent algorithm is proposed here to solve the model, and a real data example to prove its effectiveness. The rest of the paper is organized as follows. Section 2 introduces some basic concepts and theorems from uncertainty theory. In Section 3, the uncertain mathematical model for IPO problem is proposed and explained in detail. In Section 4, the techniques needed for the model are discussed. In Section 5, a numerical example and a real example are given to illustrate the effectiveness of our proposed model. A brief summary is given in Section 6.

2. Preliminary

In this section, we will introduce some fundamental concepts and results of uncertainty theory which will be used throughout the paper.

Definition 1. (Liu [12]) Let Γ be a nonempty set, and \mathcal{L} a σ -algebra over Γ ; the set function \mathcal{M} is called an uncertain measure if it satisfies the following axioms:

- (1) (Normality Axiom) $\mathcal{M}\{\Gamma\} = 1$ for the universal set Γ ;
- (2) (Duality Axiom) $\mathcal{M}\{\Lambda\} + \mathcal{M}\{\Lambda^c\} = 1$ for any event Λ ;
- (3) (Subadditivity Axiom) For every countable sequence of events $\Lambda_1, \Lambda_2, \dots$, we have

$$\mathcal{M}\left\{\bigcup_{i=1}^{\infty} \Lambda_i\right\} \leq \sum_{i=1}^{\infty} \mathcal{M}\{\Lambda_i\}.$$

The triplet $(\Gamma, \mathcal{L}, \mathcal{M})$ is called an uncertainty space.

Liu [26] defined the uncertain measure on the product σ -algebra :

- (4) (Product Axiom) Let $(\Gamma_k, \mathcal{L}_k, \mathcal{M}_k)$ be uncertainty spaces for $k=1,2,\dots$. The product uncertain measure \mathcal{M} is an uncertain measure satisfying

$$\mathcal{M}\left\{\prod_{i=1}^{\infty} \Lambda_k\right\} = \bigwedge_{k=1}^{\infty} \mathcal{M}_k\{\Lambda_k\}$$

where Λ_k are arbitrarily chosen events from \mathcal{L}_k for $k=1,2,\dots$, respectively.

The concept of uncertain variable is:

Definition 2. (Liu [12]) Let $(\Gamma, \mathcal{L}, \mathcal{M})$ be an uncertainty space. An uncertain variable is a measurable function ξ from $(\Gamma, \mathcal{L}, \mathcal{M})$ to the set of real numbers such that $\{\xi \in B\}$ is an event for any Borel set B of real numbers.

Definition 3. (Liu [12]) The uncertainty distribution Φ of an uncertain variable ξ is defined by

$$\Phi(x) = \mathcal{M}\{\xi \leq x\}$$

for any real number x .

Example 1: A zigzag uncertain variable $\xi \sim Z(a, b, c)$ has the uncertainty distribution

$$\Phi(x) = \begin{cases} 0, & \text{if } x \leq a \\ \frac{(x-a)}{2(b-a)}, & \text{if } a \leq x \leq b \\ \frac{(x+c-2b)}{2(c-b)}, & \text{if } b \leq x \leq c \\ 1, & \text{if } x \geq c. \end{cases}$$

Definition 4. (Liu [13]) An uncertainty distribution Φ is said to be regular if its inverse function $\Phi^{-1}(\alpha)$ exists and is unique for each $\alpha \in (0,1)$.

In practice, we assume that all the uncertainty distributions used in programming problems are regular. Due to the following result, we can transform an indeterminacy model to a determinacy one.

Theorem 1. (Liu [13]) Let $\xi_1, \xi_2, \dots, \xi_n$ be independent uncertain variables with regular uncertainty distributions $\Phi_1, \Phi_2, \dots, \Phi_n$, respectively, and $f: \mathfrak{R}^n \rightarrow \mathfrak{R}$, a continuous and strictly increasing function. Then the uncertain variable $\xi = f(\xi_1, \xi_2, \dots, \xi_n)$ has inverse uncertainty distribution

$$\Psi^{-1}(\alpha) = f(\Phi_1^{-1}(\alpha), \Phi_2^{-1}(\alpha), \dots, \Phi_n^{-1}(\alpha)).$$

Definition 5. (Liu [12]) The expected value of an uncertain variable ξ is defined by

$$E[\xi] = \int_0^{+\infty} \mathcal{M}\{\xi \geq x\} dx - \int_{-\infty}^0 \mathcal{M}\{\xi \leq x\} dx$$

provided that at least one of the two integrals exists.

Theorem 2. (Liu [13]) Let ξ be an uncertain variable with regular uncertainty distribution Φ . If the expected value $E[\xi]$ exists, then

$$E[\xi] = \int_0^1 \Phi^{-1}(\alpha) d\alpha.$$

Theorem 3. (Liu [13]) Let ξ and η be independent uncertain variables with finite expected values. Then for any real numbers a and b, we have

$$E[a\xi + b\eta] = aE[\xi] + bE[\eta].$$

3. IPO problem model under uncertain environment

In the IPO problem, we assume that (1) the limit values are given in advance; (2) the propagation time could be 0 in the spreading time; (3) the propagation times are uncertain variables with known uncertainty distributions. We introduce the following parameters and notations:

k : the days of propagation;

x_k : the kth day;

b_k : the limit value of propagation times;

$\xi_{x_{k-1}x_k}$: the number of uncertain propagation times in the kth day;

$\Phi(\xi)$: uncertainty distribution of $\xi_{x_{k-1}x_k}$.

We employ x and λ as the decision variables of IPO problem, where

$x = (x_1, x_2, \dots, x_n)$: integer decision vector representing n days with $1 \leq x_k \leq n$ and $x_k \neq x_{k'}$ for all $k \neq k'$, $k, k' = 2, 3, \dots, n$. Which means, the sequence x_1, x_2, \dots, x_n is a rearrangement of $\{2, 3, \dots, n\}$;

$\lambda = \{\lambda_{x_1x_2}, \lambda_{x_2x_3}, \dots, \lambda_{x_{k-1}x_k}, \dots, \lambda_{x_{n-1}x_n}\}$: 0-1 decision vector, which is:

$$\lambda_{x_{k-1}x_k} = \begin{cases} 1, & \text{if the news spread in the kth day} \\ 0, & \text{otherwise.} \end{cases}$$

According to the previous presentation and the given assumptions, the model will be proposed in an uncertain environment. We first give an example to illustrate the model.

Example 2: Assume an event happened, the experts forecast that it is likely to last for five days ($k=5$), they show the forecast value of propagation times $\xi_{x_{k-1}x_k}$ for each day, and the limit value b_k is given. If we have $x=(2,5,3,6,4)$, $\lambda = \{0,1,1,0,1\}$, that means, in the next five days, the given forecast value $\xi_{x_{k-1}x_k}$ should be scheduled as the calculated sequence so that we could get the maximized influence. The daily propagation times should be controled like this: the second day no spreading, the third day spreads control in the fifth $\xi_{x_{k-1}x_k}$ of the given forecast value, the fourth day spreads with the third forecast value, with no spread in the fifth day, and the sixth day spreads as the fourth value.

Now, let $f_k(\lambda, \xi)$ be the number of uncertain propagation times in the kth day, $k=2,3,\dots,n$, respectively. For each \$kth\$ day with $2 \leq k \leq n$, if the news spread, then we have

$$f_{x_{k-1}x_k}(\lambda, \xi) = \lambda_{x_{k-1}x_k} \xi_{x_{k-1}x_k}$$

and

$$f_{x_{k-1}x_{k+1}}(\lambda, \xi) = f_{x_{k-1}x_k}(\lambda, \xi) + \lambda_{x_kx_{k+1}} \xi_{x_kx_{k+1}}.$$

Thus, the value

$$f(\lambda, \xi) = \sum_{k=2}^n \lambda_{x_{k-1}x_k} \xi_{x_{k-1}x_k}$$

is the total time during the period of public opinion communication.

The purpose of the model is to maximize the expected value and control the trend. Thus, we have the objective function

$$E[f(\lambda, \xi)] = E\left[\sum_{k=2}^n \lambda_{x_{k-1}x_k} \xi_{x_{k-1}x_k}\right].$$

We assume that $\xi_{x_{k-1}x_k}$ has the uncertainty distribution $\Phi(\xi)$, then we obtain the inverse uncertainty distribution $\Phi_{\xi}^{-1}(\alpha)$. From theorems 2 and 3, we have

$$E[\sum_{k=2}^n \lambda_{x_{k-1}x_k} \xi_{x_{k-1}x_k}] = \sum_{k=2}^n \lambda_{x_{k-1}x_k} \int_0^1 \Phi_{\xi_{x_{k-1}x_k}}^{-1}(\alpha) d\alpha.$$

In the meantime, if we hope that the number of propagation times control in b_k in the k days ($2 \leq k \leq n$) with confidence level α , then we have the following chance constraint:

$$M\{f(\lambda, \xi) \leq b_k\} \geq \alpha.$$

If we want to maximize the impact of network public opinion, the following models are proposed:

$$\left\{ \begin{array}{l} \max E[f(\lambda, \xi)] \\ \text{subject to :} \\ \mathcal{M}\{f(\lambda, \xi) \leq b_k\} \geq \alpha, \\ \lambda \in \{0, 1\}, \\ 2 \leq x_k \leq n, \\ x_k \neq x_{k'}, \quad k \neq k', \quad k, k' = 2, 3, \dots, n \\ k = 2, 3, \dots, n, \\ x_k, k, \text{ integers.} \end{array} \right.$$

For the convenience of calculation, the above model is converted into the equivalent form:

$$\left\{ \begin{array}{l} \max \sum_{k=2}^n \lambda_{x_{k-1}x_k} \int_0^1 \Phi_{\xi_{x_{k-1}x_k}}^{-1}(\alpha) d\alpha \\ \text{subject to :} \\ \sum_{k=2}^n \lambda_{x_{k-1}x_k} \Phi_{\xi_{x_{k-1}x_k}}^{-1} \leq b_k, \\ \lambda_{x_{k-1}x_k} \in \{0, 1\}, \\ 2 \leq x_k \leq n, \\ x_k \neq x_{k'}, \quad k \neq k', \quad k, k' = 2, 3, \dots, n \\ k = 2, 3, \dots, n, \\ x_k, k, \text{ integers.} \end{array} \right.$$

where $\Phi_{\xi_{x_{k-1}x_k}}^{-1}(\alpha)$ is the inverse uncertainty distributions of $\xi_{x_{k-1}x_k}$.

4. Hybrid intelligent algorithm

In this section, we will give a practical algorithm for the uncertain IPO problem. The main idea is to apply the model to solve real cases by the given algorithm.

4.1 99-method for expected value.

The 99-method was proposed by Liu [13] on the basis of solving inverse uncertainty distribution method. This method is a good tool to solve the expected value of uncertain variables. Here, we will use it to solve the IPO problem.

From Theorem 1, we can get the the inverse uncertainty distribution of $f(\lambda, \xi)$ (Table 1):

$$\Phi_f^{-1}(\lambda, \alpha) = f(\lambda, \Phi_1^{-1}(\alpha), \Phi_2^{-1}(\alpha), \dots, \Phi_n^{-1}(\alpha)).$$

Table 1. The presentation of inverse uncertainty distribution $f(\lambda, \xi)$

α	$\Phi_f^{-1}(\lambda; \alpha)$
0.01	$f(\lambda; y_1^1, y_1^2, \dots, y_1^n)$
0.02	$f(\lambda; y_2^1, y_2^2, \dots, y_2^n)$
0.03	$f(\lambda; y_3^1, y_3^2, \dots, y_3^n)$
...	...
0.99	$f(\lambda; y_{99}^1, y_{99}^2, \dots, y_{99}^n)$

Assume $\xi_{x_{k-1}x_k}$ is an uncertain variable with the inverse uncertainty distribution $\Phi_{\xi_{x_{k-1}x_k}}^{-1}(\alpha)$,

according to the proposed model and Theorems 2 and 3, we have

$$E[\sum_{k=2}^n \lambda_{x_{k-1}x_k} \xi_{x_{k-1}x_k}] = \sum_{k=2}^n \lambda_{x_{k-1}x_k} \int_0^1 \Phi_{\xi_{x_{k-1}x_k}}^{-1}(\alpha) d\alpha$$

Thus we design the $E[f(\lambda, \xi)]$ process as follows:

Step 1 : Set $S=0, j=1$;

Step 2 : Let j take values from 0.01 to 0.99, with each increment being 0.01;

Step 3 : For each j , calculate $y_j = jf(\lambda; y_j^1, y_j^2, \dots, y_j^n)$ and $S = S + y_j$;

Step 4 : If for $j=0.99$, the value of $S + y_j$ is the expected value, stop; otherwise, turn to step 2.

4.2 Hybrid intelligent algorithm.

The concept of genetic algorithm was proposed by Holland [27] and has been applied to many optimization models. Liu [13] has embedded genetic algorithm into uncertain programming, so in order to solve the IPO problem, we will combine the 99-method with genetic algorithm. In our hybrid intelligent algorithm, the uncertain expected values are calculated by 99-method, and the final result will come out of the hybrid intelligent algorithm in order to find the optimal solution. Since the decision vectors have integer components, the integer encoding will be used for coding. Synthesizing the improved genetic algorithm and 99-method, we obtain the algorithm steps as follows:

Step 1 : Determine the population size $popsiz$ e, crossover probability P_c , mutation probability P_m and generation number N ;

Step 2 : Generate the initial $popsiz$ e chromosomes (x, y) ;

Step 3: Calculate the objective function value $f_i(x, y)$ for each chromosome (x_i, y_i) by 99-method;

Step 4 : Decide the value of v (usually $v=0.05$), calculate the fitness function g_i , select chromosomes to constitute new population --*newpop* by using roulette;

Step 5 : According to the crossover probability P_c , the crossover operation of the new population --*newpop* is carried out, and generate a new population to cover *newpop*, which is shown in FIG.1;

Step 6 : According to the mutation probability P_m , the mutation operation of the new population --*newpop* is carried out, and generate a new population to cover *newpop*;

Step 7 : Repeat Step 3 to Step 6 N times;

Step 8 : Choose the best chromosome which has the ultimate objective function value as the optimal solution.

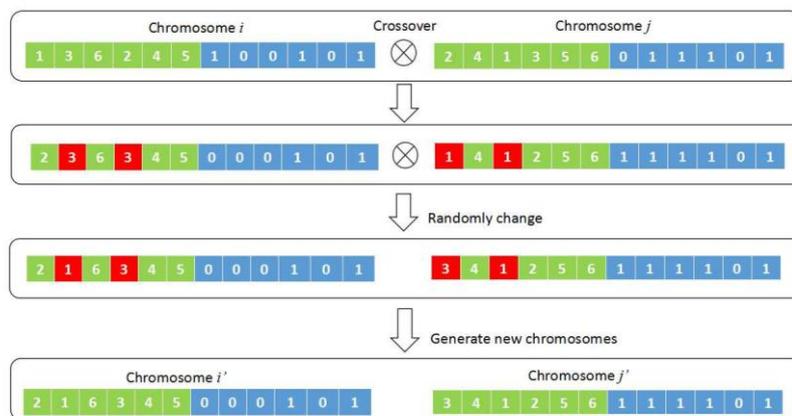


Fig. 1 An example of crossover process

5. Examples

In this section, we will verify the validity of the proposed model by a real example which happened on October 2015 in Hebei Province, China. The example about Shanhaiguan removed the 5A scenic area, and the event caused the media attention and forwarding after being exposed. However, since the key point had some deviations when the news was reported, netizens' attentiveness was not high. Hence the news keeps too long time which shown in Fig. 2.

Now, according to the propagation times on October 9th and the focus of public opinion dissemination of the subject, the possible predictive value of propagation times in the next few days is given by the experts, as they believe that the daily possible propagation times behave like zigzag uncertain variables (Table 2), where an upper limit value (Table 3) is given by the government. Here according to the proposed model, we calculate the result.

Table 2 The number of propagation times(*10)

$\xi_{x_{k-1}x_k}$	$\mathcal{Z}(a_{x_{k-1}x_k}, b_{x_{k-1}x_k}, c_{x_{k-1}x_k})$
$\xi_{x_1x_2}$	$\mathcal{Z}(1, 2, 3)$
$\xi_{x_2x_3}$	$\mathcal{Z}(40, 60, 80)$
$\xi_{x_3x_4}$	$\mathcal{Z}(4, 10, 15)$
$\xi_{x_4x_5}$	$\mathcal{Z}(320, 450, 520)$
$\xi_{x_5x_6}$	$\mathcal{Z}(20, 50, 70)$
$\xi_{x_6x_7}$	$\mathcal{Z}(50, 70, 90)$

Table 3 The upper limits

days	upper limits(*10)
2	80
3	90
4	150
5	150
6	650
7	730

Relevant calculations are carried out using Matlab tools for programming, set the population size to be 50, crossover probability P_c to be 0.5, mutation probability P_m to be 0.2, and generation number N as 1000, under the confidence level $\alpha = 0.85$. The optimal result is achieved:

$$x = (3, 2, 6, 4, 5, 7), \lambda = (1, 1, 1, 0, 1, 1).$$

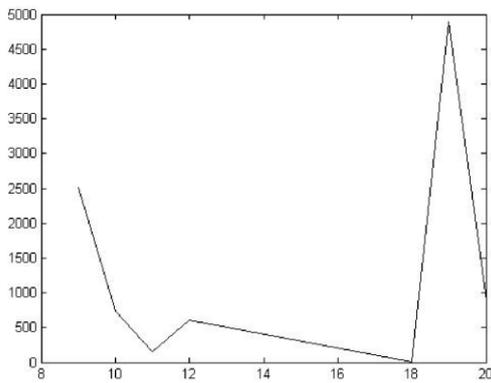


Fig. 2 The trend of Weibo attention degree

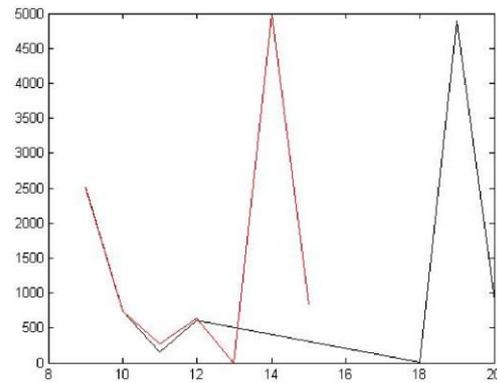


Fig. 3 Comparison chart

Therefore, under the premise of not exceeding the upper limit, according to this optimal control scheme that we propose, it can achieve the optimal effect. We compare with the actual data in FIG.3.

Usually, an internet public opinion will not last for more than 6 days. However, from the date we can see that due to lack of awareness, the event lasted as long as 12 days, but if we used the proposed IPO problem model, it would be labour saving and cost saving. Thus, we can say the optimal scheme is basically consistent with the model, and the IPO model under uncertain environment is effective.

6. Summary

In this paper, we notice the importance of uncertain factors, thus we solve the IPO problem by using uncertain variables. A novel mathematical model for IPO problem is first proposed under uncertain environment, using uncertain variables to express the uncertain factors in the model has the goal of improving the effectiveness of the model. Meanwhile, a genetic algorithm is designed for the uncertain model which is suitable for the IPO control problem. At last, the IPO problem model is applied to a real event, the result shows a good performance.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (No.2016MS65), and the National Natural Science Foundation of China (No.71671064).

References

- [1]. A Sudbury. The proportion of the population never bearing a rumour. Journal of Applied Probability. Vol.22(1985)No.2, p.443-446.

- [2]. Y Lan, X Dong. Research on Micro-blog Public Opinion Information Interaction Model Under the Background of Big Data. *New Technology of Library & Information Service*. Vol.258(2015)No.5, p.24-33.
- [3]. H Sun. Research on Network Public Opinion Propagation Model (MS., Shandong University, China 2014). p.45-57.
- [4]. D Zanetee. Criticality of rumor propagation on small world networks. *Statistical Mechanics & its Applications*. 2008, p.1-10.
- [5]. D Zanetee. Dynamics of rumor propagation on small world networks. *Physical Review E*. Vol.65(2001)No.1, p.110-126.
- [6]. Z Pan, X Wang, X Li. Simulation Investigation on Rumor Spreading on Scale-free Network with Tunable Clustering. *Journal of System Simulation*. Vol.18(2006)No.8, p.2346-2348.
- [7]. S Fortunato. Damage spreading and opinion dynamics on scale-free networks. *Physica A: Statistical Mechanics & its Applications*. Vol.348(2005), p.683-690.
- [8]. M He, D Zhang, H Wang, et al. Public opinion evolution model with the variable topology structure based on scale free network. *Acta Physica Sinica*. Vol.59(2010)No.8, p.5175-5181.
- [9]. Y Moreno, M Nekovee, A Pacheco. Dynamics of rumor spreading in complex networks. *Physical Review E Statistical Nonlinear & Soft Matter Physics*. Vol.69(2004)No.2, p.279-307.
- [10]. Y Wang, K You, M Wang, et al. Model of network community public opinion spread based on game theory. *Application Research of Computers*. Vol.30(2013)No.8, p.2480-2482.
- [11]. C Liu, X Hu, G Si, et al. Public Opinion Propagation Model Based on Small World Networks. *Journal of System Simulation*. Vol.18(2006) No.12, p.3608-3610.
- [12]. B Liu. *Uncertainty Theory*, 2nd ed.. Springer-Verlag, Berlin, 2007.
- [13]. B Liu. *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty*. Springer-Verlag, Berlin, 2010.
- [14]. X Chen, B Liu. Existence and uniqueness theorem for uncertain differential equations. *Fuzzy Optimization & Decision Making*. Vol.9(2010)No.1, p.69-81.
- [15]. K Yao. Extreme values and integral of solution of uncertain differential equation. *Journal of Uncertainty Analysis & Applications*. Vol.1(2013)No.1, p.1-21.
- [16]. B Liu. Fuzzy process, hybrid process and uncertain process. *Journal of Uncertain Systems*. Vol.2(2008)No.1, p.3-16.
- [17]. B Liu. *Theory and Practice of Uncertain Programming*, 2nd ed.. Springer-Verlag, Berlin, 2009.
- [18]. X Li, Z Qin. Interval portfolio selection models within the framework of uncertainty theory. *Economic Modelling*. Vol.41(2014), p.338-344.
- [19]. B Liu. Uncertain set theory and uncertain inference rule with application to uncertain control. *Journal of Uncertain Systems*. Vol.4(2010)No.2, p.83-98.
- [20]. K Yao. Inclusion relationship of uncertain sets. *Journal of Uncertainty Analysis & Applications*. Vol.3(2015)No.1, p.1-5.
- [21]. X Gao, L Jia, S Kar. A new definition of cross-entropy for uncertain variables. *Soft Computing*. DOI:10.1007/s00500-017-2534-6.
- [22]. J Peng, B Zhang, S Li. Towards uncertain network optimization. *Journal of Uncertainty Analysis & Applications*. Vol.3(2015)No.1, p.1-19.
- [23]. X Gao, L Jia. Degree-constrained minimum spanning tree problem with uncertain edge weights. *Applied Soft Computing*. DOI: 10.1016/j.asoc.2016.07.054.
- [24]. X Gao, Y Gao, D Ralescu. On Liu's Inference Rule for Uncertain Systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. Vol.18(2010)No.1, p.1-11.
- [25]. C Su, J Peng, S Li. The internet public opinion propagation model via uncertain differential equation. *Systems Engineering-Theory & Practice*. Vol.35(2015)No.12, p.3201-3209.
- [26]. B Liu. Some research problems in uncertainty theory. *Journal of Uncertain Systems*. Vol.3(2009)No.1, p.3-10.
- [27]. J Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor. 1975.