

# Gear Fault Diagnosis Correlation Analysis Based on Probability Box Theory

Liu An<sup>1</sup>

Faculty of Mechanical and Electrical Engineering  
Kunming University of Science and Technology  
Kunming, China  
[liuanmike@sina.com](mailto:liuanmike@sina.com)

Du Yi<sup>2</sup>

City College  
Kunming University of Science and Technology  
Kunming, China  
[duyi215@hotmail.com](mailto:duyi215@hotmail.com)

He Wei<sup>3</sup>

Faculty of Mechanical and Electrical Engineering  
Kunming University of Science and Technology  
Kunming, China  
[Hw\\_big@126.com](mailto:Hw_big@126.com)

Ding Jiaman<sup>4</sup>

College of Information Engineering and  
Automationline  
Kunming University of Science and Technology  
Kunming, China  
[tjom2008@163.com](mailto:tjom2008@163.com)

**Abstract**—Measuring signal in gearbox fault diagnosis low signal-to-noise ratio, high frequency of characteristic signal and noise reduction was difficult. And general gear fault diagnosis method was hard to solve the space-time registration problem of multiple source information. So probability box theory and information fusion methods were applied in this paper, which introduced the basic development process of the probability boxes and the methods of information fusion. However, independent irrelevance of measured signal was assumed before information fusion, which was unreasonable. For this reason, classification strategy and pattern of correlation based on probability box theory was put forward. The experiment proved that correlation classification for multiple source information during gear fault diagnosis improved fusion efficiency and science, can better met the authenticity of the space-time registration problem of multiple source information. which developed a new approach to gear fault diagnosis or bear fault diagnosis research to solve the space-time registration problem of multiple source information.

**Keywords**- *gear fault;probability box theory;space-time registration ;information fusion; correlation;*

## I. INTRODUCTION

Gear transmission is commonly used in the machinery transmission ways. In many cases, the gear failure is an important factor of inducing mechanical failure. So the gear fault diagnosis is very important and practical significance.

Gear box vibration noise is big and the environment is poor in the practical work, the extracted signal to noise ratio of gear vibration signal is low, especially the fault occurred in early. The fault feature signals submerged in noise, the classical Fourier spectrum analysis method and commonly used wavelet transform were difficult to extract the fault frequency feature [1-3]. At the same time, it is difficult to solve the space-time registration problem based on multi-source information.

Probability boxes theory provides a new research way to solve the above two problem. In 1966 Moore[4] firstly put forward "cognitive uncertainty" by using interval to definite

it purely, named "Type I uncertainty". "Accidental Uncertainty" usually expressed by probability theory. In 1967 Dempster [5] proposed the expression of probability upper and lower bound that is based on multivariate mapping relation. Shafer [6] put forward the famous Dempster Shafer evidence theory in 1986, the theory takes basic probability distribution function as the core, with belief functions and likelihood function contain uncertainty problem, which meet more weaker condition than Bayesian inference. In 1990, Williamson [7] put forward the expression of correlation and convolution of interval probability algorithm in "Type I uncertainty" problem. American Sandia National laboratory in 2003 [8] promoted the project of cognitive uncertainty, the laboratory combined DS evidence theory with probability analysis, put forward at the same time the express of "Type I uncertainty" and "Type II" uncertainty" based on Dempster Shafer Structure (DSS) theory of probability box theory. Probability box theory applied successively involves fault system failure probability assessment [9], the dynamic responses of the vibration system uncertainty evaluation [10], the uncertainty expression of climate change [11] and the application is being extended.

A probability box is the class of distribution function  $\underline{F}(x)$  and  $\overline{F}(x)$  such that  $\underline{F}(x) \leq F(x) \leq \overline{F}(x)$  for all  $x$  value. Probability boxes thus express interval-like uncertainty about a distribution function. Probability bounds analysis was the collection of methods and algorithms that were used to do calculations with, and made inferences from, probability boxes. The probability boxes theory can take subjective and objective uncertainties into account and make up the defect of discarding rich statistical information due to feature extraction process. The raw data is firstly converted into a probability box and then probability boxes are fused. Using probability box as the space-time registration framework can solve the second problem [12].

General speaking, the information obtained from a single sensor is very limited, and it just feedback the partial signal of environment characteristic. Multiple sensors can get a lot of information from the whole space angles, but there is a redundancy and conflict between information. Therefore, the correlation analysis can eliminate the false and retain the true, reflecting the system state fully.

Correlation classification strategy and pattern were summarized. The experiment proved the application value of perfect correlation and positive correlation. At last, the fused information was integrated to probability box, which laid the foundation for gear fault pattern recognition based on probability box theory.

## II. RESEARCH ON INFORMATION FUSION BASED ON PROBABILITY BOX THEORY DURING GEAR FAULT DIAGNOSIS

Information fusion is a theory processing multiple source information. That is to say, comprehensive treatment the information of different time and space, explore a variety of physical quantities by using multisensory, grading processing multiple source and multiple data, determine the system state accurately and timely, provide the right judgment of system fault and fault mode, and analyze the relationship between statement(fault)、phenomena and cause[13].

Based on different actual demand and feasibility, a lot of information fusion models were put forward [14]. For classification model, which can be divided into two broad categories. One was that the external characteristics model based on the system, such as function model, structure model, the position fusion model, input and output model and so on. The other was that the own characteristic model based on fusion algorithm such as layered fusion model, convergence properties and so on.

When the model is established, the algorithm is the key for information fusion. Measured signal signal-noise is low, feature signal frequency is high, transfer path is complex, noise reduction is difficult, those are characteristic of gear fault diagnosis. Probability box theory has great advantage of processing original signal's containment and uncertainty. One is considering subjective and objective factors, which reflects the comprehensive uncertain veritably. The other is that it can solve the multiple source space-time registration problem.

However, independent irrelevance of measured signal was assumed before information fusion, which was unreasonable. Evidence theory keep conservatism when solve contradictory data, so some scholars put forward different information fusion algorithm [15-17]. Measured on the same target of multiple source information exist the redundancy and complementarity of space and time, which there is a certain correlation. So analyzing the multiple source correlation contribute to grasp the internal connection between the information, which develop a new method based on probability box theory information fusion algorithm.

## III. CORRELATION CLASSIFICATION BASED ON PROBABILITY BOX THEORY

Based on probability theory, we defined the correlation between event A and event B like this [18].

Assuming that the small circle denotes the event A probability interval, the great circle denotes the event B probability interval. Correlation classification between event A and B as follow Fig .1:

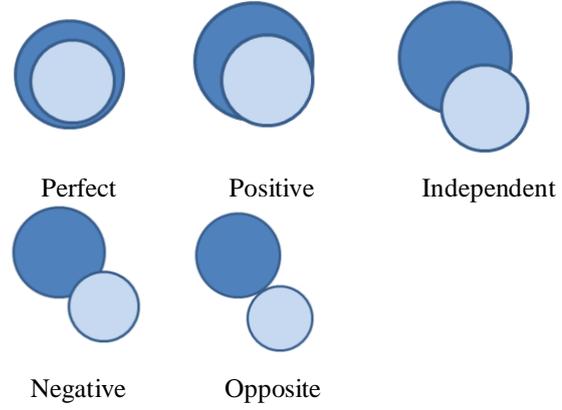


Figure 1: Correlation classification between two events depicted in Venn diagrams.

Perfect correlation denotes happened in a small circle of events will occur in the great circle.

Opposite correlation denotes happened in a small circle of events will not occur in the great circle.

Independent correlation denotes intersection is small between event A and B probability interval, or considered intersection is empty set, then define event A and B are independent uncorrelated.

Define the case between perfect correlation and independent uncorrelated as positive correlation.

Define the case between opposite correlation and independent uncorrelated as negative correlation.

The formulas for probabilities of conjunction and disjunction under perfect dependence are:

$$P(A \wedge B) = and_{\text{perfect}}(a, b) = \min(a, b) \quad (1)$$

$$P(A \vee B) = or_{\text{perfect}}(a, b) = \max(a, b) \quad (2)$$

The formulas for probabilities of conjunction and disjunction under opposite dependence are:

$$P(A \wedge B) = and_{\text{opposite}}(a, b) = \max(a + b - 1, 0) \quad (3)$$

$$P(A \vee B) = or_{\text{opposite}}(a, b) = \min(1, a + b) \quad (4)$$

The formulas for probabilities of conjunction and disjunction under positive dependence are:

$$P(A \wedge B) = and_{\text{positive}}(a, b) = [ab, \min(a, b)] \quad (5)$$

$$P(A \vee B) = or_{\text{positive}}(a, b) = [1 - (1-a)(1-b), \max(a, b)] \quad (6)$$

The formulas for probabilities of conjunction and disjunction under negative dependence are:

$$P(A \wedge B) = and_{\text{negative}}(a, b) = [\max(a+b-1, 0), ab] \quad (7)$$

$$P(A \vee B) = or_{\text{negative}}(a, b) = [1 - (1-a)(1-b), \min(1, a+b)] \quad (8)$$

#### IV. CORRELATION EXPRESSION FORM BASED PROBABILITY BOX THEORY

Assume that E denotes the expectation and V denotes the variance, r denotes the correlation between two events, as formula 9

$$r = \frac{E(AB) - E(A)E(B)}{\sqrt{V(A)}\sqrt{V(B)}} \quad (9)$$

Assume that the expectation of an indicator function for an event A is the probability of the event P(A). Formula 9 can evolve into Formula 10, as follow

$$r = \frac{P(A \wedge B) - P(A)P(B)}{\sqrt{P(A)(1-P(A))}\sqrt{P(B)(1-P(B))}} \quad (10)$$

Apply r to Lucas correlation model, get the formula 11:

$$P(A \wedge B) = and_{\text{Lucas}}(a, b, r) = ab + r\sqrt{a(1-a)b(1-b)} \quad (11)$$

Where  $a = P(A)$   $b = P(B)$

Correlation exist uncertainty as well, using probability box theory express the uncertainty, formula 12 defined

$$\underline{r} = \frac{\max(a+b-1, 0)}{\sqrt{a(1-a)b(1-b)}} \quad (12)$$

$$\bar{r} = \frac{\min(a, b) - ab}{\sqrt{a(1-a)b(1-b)}} \quad (13)$$

Whereby,  $\underline{r}$  denotes correlation factor r estimate lower limit,  $\bar{r}$  denotes correlation factor r estimate upper limit, correlation r to be no smaller than  $\underline{r}$  and no larger than  $\bar{r}$ .

#### V. INFORMATION FUSION CORRELATION EXPERIMENT BASES ON PROBABILITY BOX THEORY

In order to testify the fusion difference under the condition of different correlation based on probability box theory, we assumed that two gear fault experiment

information source obeyed exponential distribution, Dempster Shafer Structure of the exponential distribution parameter  $\lambda$  as follow:

$$DSS_1 = [2, 3, 1], \quad DSS_2 = [2.1, 2.9, 1]$$

Contrasting two information source, the Dempster Shafer Structure of the exponential distribution parameter  $\lambda$  was very close, which belong to perfect correlation. Under the independent uncorrelated assume, two probability box fusion result using DS evidence theory is Fig .2. In order to clearly express purpose, it cut out the back of the data of more than 10 in the horizontal direction. Under the condition of perfect correlation, taking Lucas correlation model's correlation factor  $[\underline{r}, \bar{r}]$  inset DS evidence fusion algorithm from formula 12, the result was Fig .2.

Contrasting two tables, it indicated that the fusion result under the condition of the perfect correlation was more reasonable than under the assumption of independent uncorrelated. The information source fusion result under the perfect correlation condition were no different than before, but the result of the Fig .2(a) was less than satisfactory distinctly. The paper also built some models in view of the correlation between other conditions, the result improved contrasting the default fusion algorithm.

Numerical quantitative analysis to the results of Fig .2 and Fig .3, by using the method of definite integral area. Assumed that the probability area of the p-box 1、 p-box 2 and the fusion between p-box 1 and p-box 2 respectively were: A1, A2, A3. Defined the fusion overlap factor  $\tau$  was formula 13:

$$\tau = \frac{2A_3}{A_1 + A_2} \times 100\% \quad (13)$$

The overlap factor numerical value of the default DS fusion algorithm and the modified DS fusion algorithm under the perfect correlation condition respectively were 59.7% and 97.6%.

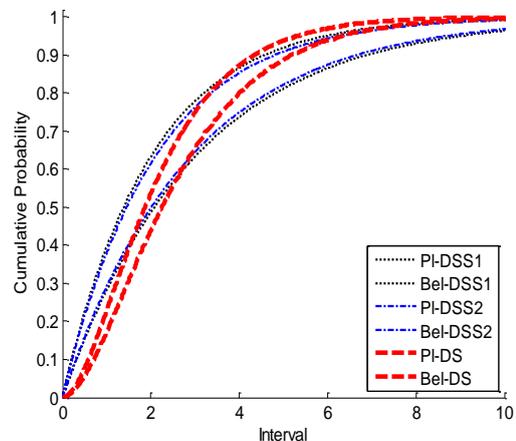


Figure 2. the analog signals' DS evidence fusion result comparison under independent correlation condition

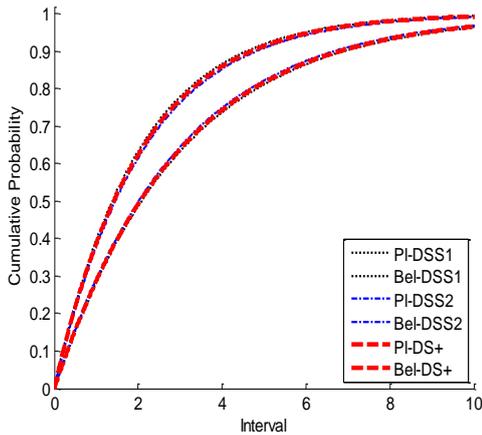


Figure 3. the analog signals' DS evidence fusion result comparison under positive correlation condition

On the gear fault test bench, the paper studied the gear crack, lay out the sensor as follow:

Two acceleration sensors were distributed in 2mm distance on the bearing seat, which got perfect correlation.

On the top and the side of the bearing seat distributed an acceleration respectively, which got positive correlation.

It got vibration acceleration amplitude signals from gear crack acceleration sensor, then the singular value denoising processing, and according to the sampling frequency grouping processing, getting the DSS data of each group. According to DSS value, built two sensor p-boxes owing to the different layout. The paper computed DS fusion algorithm under the independent uncorrelated condition and the modified DS fusion algorithm based on Lucas correlation model, the overlap factor value were presented in Table 1.

TABLE I. GEAR CRAKE FUSION RESULT CONTRAST UNDER DIFFERENT COLLELATION CONDITON

Experiment type	Perfect correlation experiment		Positive correlation experiment	
	Independent	Perfect	Independent	Positive
$\tau$	80%	100%	57.8%	97.8%

The Table 1 indicated that gear crack fault result was similar between perfect correlation experiment and analog signal. The experiment proved that the modified DS fusion algorithm based on Lucas correlation model was effective for gear crack fault diagnosis. The paper were made a similar experiment for other gear wear and plastic deformation, the result was similar.

## VI. CONCLUSION

The paper expounded the classification strategy and pattern of information fusion correlation based on probability box theory, the experiment proved that the data fusion difference between different correlation. On the

gear fault text bench, it got perfect correlation data and positive correlation data .The paper computed DS fusion algorithm under the independent uncorrelated condition and the modified DS fusion algorithm based on Lucas correlation model. The experiment proved that the latter was effective for gear crack fault diagnosis. Therefore, gear fault diagnosis correlation classification for multiple source information improved fusion efficiency, which laid a foundation for pattern recognition.

## ACKNOWLEDGMENT

The corresponding author of this paper is Du Yi. This project is supported by National Natural Science Foundation of China (Grant No.51365020, 51369012), Yunnan Natural Science Foundation (Grant No.2013FZ020)

## REFERENCES

- [1]Yuan Jiasheng ,Feng Zhihua.Fault Diagnosis Research Based on Correlation and Wavelet for Gears[J],Tracsaction of the Chinese Society for Agricultural Machinery.2007,38(8),120-123
- [2]Xu Yong-gang, Meng Zhi-peng, Lu Ming, Fu Sheng.Gear fault diagnosis based on dual-tree complex wavelet transform and singular value difference spectrum[J],Journal of Vibration and Shock,vol.33 No.1 2014
- [3]Lei Yaguo, He Zhengjia , Lin Jing , Han Dong ,Kong Detong. Research advances of fault diagnosis technique for planetary gearboxes[J],Journal of Mechanical engineering, Vol.47 No.19,Oct.2011
- [4]Moore.Interval Analysis.New Jersey,1996.
- [5]Dempster. Upper and lower probabilities induced by a multi-valued mapping. Annals of Mathematical Statistics, 1967, 38: 325-339
- [6]Shafer. The combination of evidence. International Journal of Intelligent Systems, 1986, 1: 155-179
- [7]Williamson and Downs. Probabilistic arithmetic I: numerical methods forcalculating convolutions and dependency bounds. International Journal of Approximate Reasoning, 1990, 4:89-158
- [8]Ferson. Constructing Probability Boxes and Dempster-Shafer Structures.Sandia National Laboratories, 2003
- [9]Shafer. A Mathematical Theory of Evidence. Princeton:Princeton University Press, 1976
- [10]Fulvio Tonon. Using random set theory to propagate epistemic uncertainty through a mechanical system. Reliability Engineering and System Safety, 2004,85:169-181
- [11]Kriegler, Held. Utilizing belief functions for the estimation of future climate change. International Journal of Approximate Reasoning, 2005, 39:185-209
- [12]Du Yi . Fault diagnosis method of roller bearing based on probability box theory and information fusion [D].Kunming,Kunming university of science and technology,2003
- [13]Xiong Jun , Li Fengying,Shen Yudi.Neural network information fusion-Fuzzy inference theory and application.Beijing:National Defence Industry Press,2002
- [14]Chen Jin.Mechanical equipment fault diagnosis technology and its application.Shanghai:Shanghai higher education electronic audio and video publishing house,2003
- [15] Ferson, Kreinovich, Hajagos. Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty, Sandia National Laboratories,Albuquerque, NM, 2007
- [16] Ronald R. Yager. Uncertainty Representation Using Fuzzy Measures. IEEE transaction on systems, 2002, 32(1): 13-20
- [17] Frank, Nelsen and Schweizer. Best-possible bounds for the distribution of a sum - a problem of Kolmogorov. Probability Theory and Related Fields, 1987, 74:199-211
- [18]Scott Ferson, Janos Hajagos,Daniel Berleant,Jianzhong Zhang,W.Troy Tucker,Lev Ginzburg and William Oberkampf.Dependence in Dempster-Shafer theory and probability bounds analysis,2004