The Application of 3D Fruit Fly Optimization Algorithm to the Classification of Keywords Data of Macau's International Relations

Shianghau WU[a, 1], Haoyu SHI[b]
[a]School of Business
[b]Institute for Social and Cultural Research
Macau University of Science and Technology
Macau, China
[1]hwu@must.edu.mo.

ABSTRACT: *In recent years, Macau has actively participated in international organizations according to the Macau Basic Law. After fifteen years of regaining sovereignty by China, Macau has faced some internal and external challenges. In order to comprehend the new policy directions regarding Macau's participation in international organizations, this paper made the depth interviews and applied the text mining analysis to the interviews responses. Through applying the 3D fruit fly optimization algorithm (3D FOA)and our modified 3D FOA model to make the classification analysis of interviewees' keywords. The study concluded the modified 3D FOA model has the better performance, and the keypoints for our interviewees regarding Macau's participation in international organizations include the Basic Law's regularization and Macau's participation in international economic organization.*

Keywords: Text Mining, Wavelet Transformation, Classification, 3D Fruit Fly Optimization Algorithm

Received: 10 October 2015, Revised 12 November 2015, Accepted 17 November 2015

1. Introduction

Macau is the special administrative region of China. Some scholars referred to Macau's international participation as the example of "paradiplomacy". Paradiplomacy is the emerging policy capacity of sub-state entities which can be enjoyed by both states and sub-state entities [1]. Some scholars think Macau's paradiplomacy as the outreach of non-sovereign jurisdictions to actors beyond their borders and frontiers of their metropolitan relationships or claimant states. The power of the sub-states in international relations was decided by the sovereign state [2]. In recent years, Macau's international participation has become the research focus of international relations researchers. In order to comprehend the new policy directions regarding Macau's participation in international organizations, this paper made the depth interviews on two renowned specialists in politics and economy fields and applied the text mining analysis to the interviewees' responses.

1.1 Methodology

1.1.1 Text Mining

Text mining is one of the data mining methods, which learn from samples of past experience. In the text mining method, the text will be processes and transformed into a numerical representation.

According to the recent literature of text mining, the research focuses of text mining can be divided as the application of text

mining to different areas and the new design of text mining algorithm and software.

As for the text mining application research, the text mining method is widely applied to information management on websites, biological data and customer relationship management.

Other applications mainly relate to library management and literature critique. For example, Clement (2008) used the text mining method to analyze Gertrude Stein's "The Making of Americans" and has different conclusions from that of other literature critics [3]. And Trappey (2007) applied the text mining method to analyze patent brochures [4].

As for the new design of text mining algorithm and software research, Schmidt (2010) used the text mining method to analyze the marketing-oriented textual information[5]. Esuli and Sebastiani (2010) described an industrial-strength software system for automatically coding open-ended survey responses and compared the accuracy of the software against the accuracy of human coders [6]. Lee and Bradlow (2011) described a new text mining algorithm for analyzing online customer reviews to facilitate the analysis of market structure[7].

1.2. Depth Interviews

In order to explore more thorough results, the study made depth interviews on two Macau's renowned specialists in the fields of politics and economy. The aim was to comprehend the Macau's problems and challenges on the international participation and conclude Macau's new policy directions. The interviews were taken place respectively on the first week of January 2015. Each interview took about forty-five minutes. The depth interview questions were made according to the news reports and research papers related to Macau's participation in international organizations.

The depth interview questions were listed as follows:

- What is the possible development for Macau to participate in international organizations?
- According to your professional background, what kinds of international organizations are suitable for Macau's participation?
- Macau joined the Asia Pacific anti-money laundry organization. Why does Macau still link to money laundry according to some international press reports?
- What are your viewpoints about the cooperation between Macau's civil organizations and its international counterparts?
- What are the related measures for Macau Special Administrative government to handle the international affairs?
- What are your views regarding Macau's roles in international organizations?
- Can you give some examples about Macau's participation in international organizations?
- Can you make the comparison with Macau's participation in international organizations before and after 1999?
- In your opinion, what kinds of international organizations or conventions are important for Macau's participation?
- What kinds of efforts should Macau make for more roles in international organizations?
- Do you think it is important to encourage Macau citizens to participate in international affairs?
- Do you think Chinese central government's "regional cooperation" concept contradict to the Macau's internationalization?

The study assorted the depth interview response data in order to facilitate the application to the text mining method.

1.3. Research Design

The text mining method of the study is implemented by means of the text processing function of the R software and its application packages. The study uses the text mining application ("tm") package of R software to analyze the depth interview text data. The text mining method of the study is implemented by means of the text mining package (tm) and "Rwordseg" package of the R software[8]. The text mining package follows the procedures as follows[9]:

Parsing: The text and the structure were extracted and represented in a data structure. N Chars Filter: The node filters terms consisting of less than a specified number.

In the study, the specified number is set as four in order to filter out definite and indefinite articles (e.g., a, an, the).

Number Filter: The node filters all terms consisting of numbers only.

The Punctuation Erasure: The node removes all punctuation marks.

Stop Words Filter: The node filters all stop words.

The study also calculated the term frequency (tf), and the inverse document frequency (idf) to choose the major keywords. The study got major keywords from the contents of two interview responses including: "Basic Law", "central government", "APEC", "service industry", "regularize", "financial industry", "productivity", and "diversification".

The study got fifteen keywords frequency data and categorized the keywords data from the first interviewee as "Type 0" and keywords data from the second interviewee as "Type 1".

1.3.1. Wavelet Transformation

The wavelet transformation is a tool that separates data, functions, or operators into different frequency components and then explores each component with a resolution matched to its scale. Wavelet transformations were developed to express the frequency domain and the time locality of an input function. The fact that wavelets capture the temporal nature of the data is quite essential [10]

The study used the "WaveThresh" package of R language [11] to make the wavelet transformation. The wavelet transformation enables us to find the independent coefficients using either Normalized Elliptic Fourier or DiscreteWavelet, which can facilitate further data handling. The study used the original keywords idf data to make the wavelet transformation. The R commands were as follows.

```
program WaveThresh
{Let (y1, . . . ,yn )as the keywords idf data to become input . }
Begin :
ywdS <-wd ( y, filter.number = 1, family = "DaubExPhase", type = " station")
accessD (ywd, level = 2)
end
```

According to the R commands, the study got finest-scale coefficients. And the wavelet was the Daubechies wavelets as defaults[12].

2. 3D Fruit Fly Optimization Algorithm

The Fruit Fly Optimization Algorithm (FOA) is a new optimization method developed by Pan (2012)[13]. The FOA was based on the sniffing behavior of fruit flies. The FOA program can be applied to gray system, data mining and neural networks. It assumed the fruit fires group located in a random space and found the food's location by sniffing [14]. When the fruit fries reached the food's location, the swarm of fruit flies will fly to the location directly by sense vision (sniffing) and the location would be confirmed. The study used the 3D FOA program to make the classification based on the improvement of FOA in

order to have the better performance. The basic step for 3D FOA calculation was as follows [15] [16], Step 1. The method set the fruit fly group's location by random (X-axis, Y-axis, Z-axis).

Step 2. The method endowed individual fruit fly with random value for locating food location(X,Y, Z).

X= X-axis+ Random Value;

Y= Y-axis+ Random Value;

Z= Z-axis+ Random Value.

Step 3. The method calculated the personal fruit fly's distance from the original point (0,0), and found the S value of density, this value equals to the inverse value of distance. $\text{Dist} = \sqrt{x^2 + y^2 + z^2}; s = \frac{1}{dist}$

Step 4. The method induced the smell density Value S into the density judge function, and find out the smell density value of personal fruit fly at confirmed location.
Smell = Function(S).

Step 5. The method repeated the step two to step four, calculating the smell density of all the fruit flies in the swarm, and found out the fruit fly with largest and lowest smell density value.

Step 6. The method recorded the smell density value and location (X,Y,Z) of the best fruit fly, and the flying routine of the group fly to that location.

Step 7. The method made the iterative optimization, and repeated step two to step five, and checked whether the smell density value was better than the previous one, if yes, carried on the step six, or carry on the step seven, until match the max iterations, then the calculating finished.

The study attempted to comprehend the relationship of each keyword in each analyzed document. In order to achieve the goal, the study stipulated the Keywords Frequency Composite Function (KFCF), which intended to explore the weighting of each keyword in each document. The study makes the Composite Keywords Frequency Index as Z, which represents the type of each keyword document. The inverse document frequency of each keyword in each document is represented as Xi. The coefficient of each keyword is ai. The KFCF was set as follows [17],

$$Z=\sum_{i=1}^{n} a_i X_i \quad (1)$$

The study used the Fruit Fly Optimization Algorithm to get the coefficients of each keyword by the optimization process. The study attempted to fit the Keywords Frequency Composite Function by FOA optimization. The iterations of FOA were set as 100, and the size of fruit flies was set as 8, which meant each fly was designated for each keyword.

The study randomly chose one-eighth of data as the test data, and the other data as the training data. The iterative optimization process was depicted as Figure 1.

The searching route of fruit flies was depicted as Figure 2.

The study also used the ROC (Receiver Operating Characteristic) curve to determine whether the model is the suitable model. The ROC curve plots the true positive rate against the false positive rate. The method is to consider the square measures of areas under the ROC curves. If the square measure approaches to 0.5, it would be the less corresponding model. If the square measure equals to 1, it would bet the model with perfect accuracy. According to the calculation, the square measure of the area under the ROC curve was 0.8. The ROC curve of the 3D FOA model in the study is shown on Fig.3.

The error matrix of 3D FOA classification was shown as Table 1. According to Table 1, the overall error percentage of 3D FOA was 7.14%. The study got the KFCF according to the 3D-FOA as follows:

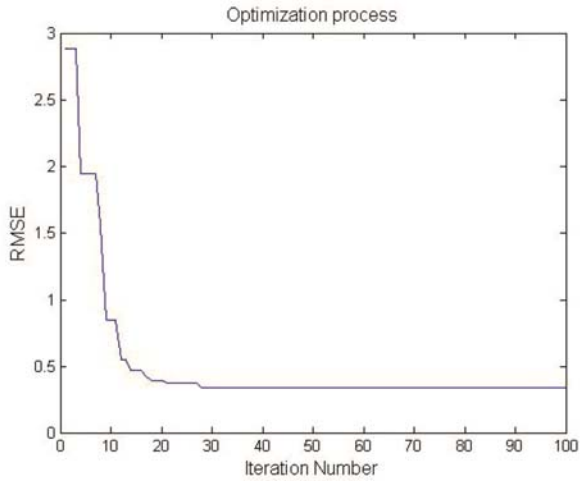

Figure 1. Iterative optimization process of 3D FOA

Z =0:2483APEC + 0.2453*ServiceIndustry+0:2649regularization +0.2438* f inancialindustry +0:2371productivity +0.2363*centralgovernment+0:2263diversification +0.3147*BasicLaw

Observed	Predicted		
	Type No.0	Type No.1	Percentage Correct
Type No. 0	10	0	0
Type No. 1	1	3	75
Overall Error Percentage	7.14%		

Table 1. Error matrix for the 3D FOA Classification Model

3. Modified 3D FOA Classification

The study also used the modified 3D Fruit Fly Optimization Algorithm to induce the coefficients of KFCF. The goal of using the Modified 3D FOA is to avoid the local extremum problem of optimization. The study improved a modified 3D FOA originated by Li(2012)[18].

The basic step of modified 3D FOA program was as follows,

Step 1. The method set the fruit fly group's location by random (X-axis, Y-axis, Z-axis).

Step 2. The method endowed individual fruit fly with random value for locating food location(X,Y, Z).

X= X-axis+ Random Value;

Y= Y-axis+ Random Value;

Z= Z-axis+ Random Value.

Step 3. The method calculated the personal fruit fly's distance from the original point (0,0), and found the S value of density, this value equals to the inverse value of distance.

$\text{Dist} = \sqrt{x^2+y^2+z^2}; s = \frac{1}{dist} + \Delta; \Delta = dist * (0.5 - \delta) * e^g; 0 \leq \delta \leq 1;$ g is the square value of RMSE between KFCF output value and the target value.

Step 4. The method induced the smell density Value S into the density judge function, and find out the smell density value of

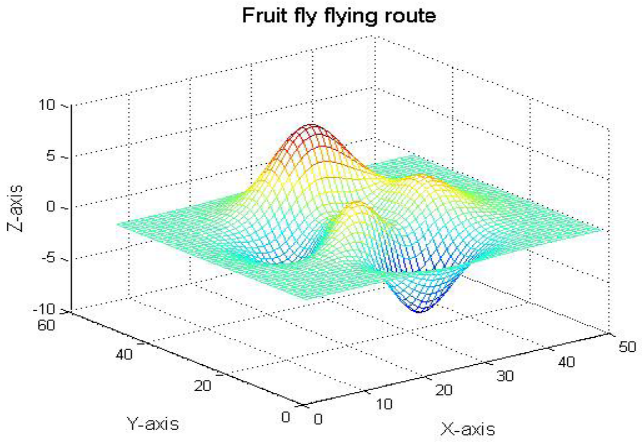

Figure 2. The Searching Route for Fruit Flies of 3D FOA

personal fruit fly at confirmed location.
Smell=Function(S).

Step 5. The method repeated the step two to step four, calculating the smell density of all the fruit flies in the swarm, and found out the fruit fly with largest and lowest smell density value.

Step 6. The method recorded the smell density value and location (X,Y,Z) of the best fruit fly, and the flying routine of the group fly to that location.

Step 7. The method made the iterative optimization, and repeated step two to step five, and checked whether the smell density value was better than the previous one, if yes, carried on the step six, or carry on the step seven, until match the max iterations, then the calculating finished.

The study used the modified Fruit Fly Optimization Algorithm to get the coefficients of each keyword by the optimization process. The study attempted to fit the Keywords Frequency Composite Function by modified FOA optimization. The iterations of FOA were set as 100, and the size of fruit flies was set as 8, which meant each fly was designated for each keyword.The study randomly chose one-eighth of data as the test data, and the other data as the training data. The iterative optimization process was depicted as Figure 4.The searching route of fruit flies was depicted as Figure 5.

The error matrix of modified 3D FOA classification was shown as Table 2. According to Table 2, the overall error percentage of modified 3D FOA was 7.14%. According to the calculation, the square measure of the area under the ROC curve was 0.8.

The ROC curve of the modified 3D FOA model in the study is shown on Figure 6.

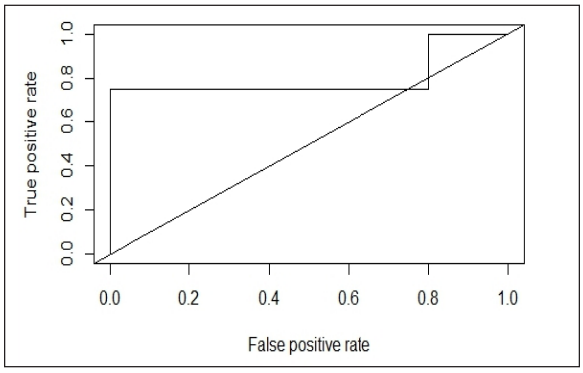

Figure 3. ROC Curve of 3D FOA Classification

Observed	Predicted		
	Type No.0	Type No.1	Percentage Correct
Type No. 0	10	0	0
Type No. 1	1	3	75
Overall Error Percentage	7.14%		

Table 2. Error matrix for the 3D FOA Classification Model

The study got the KFCF according to the modified 3D-FOA as follows:

Z =0:4119APEC + 0.4050*ServiceIndustry+0:4231regularization +0.4022* f inancialindustry
+0:3943productivity +0.3851*centralgovernment+0:3752diversification +0.5052*BasicLaw

4. Discussion

Prior studies have focused on the comparison of the calculation results between improved FOA and original FOA. Mousavi et.al. (2015), for example, propsed an improved fruit fly optimization algorithm (IFOA) to solve the fuzzy questions and made the comparison among the calculation results of IFOA and other algorithms[19].

However, these studies did not relate to the classification of text mining results. The study made the comparison of the original and improved 3D FOA classification. The study provides the results as follows.

First, the study finds that the improved 3D FOA classification has the lower RMSE(Root Mean Square Error)in the iterative optimization process. According to Figure 1 and Figure 4, the RMSE of improved 3D FOA classification is 0.3369 after 100 iterations. It means the improved 3D FOA has better the performance of optimization.

Second, we can find that the square measure of the area below the ROC curve of the two classification models is 0.8, and the overall error percentage of the two models is 7.14%. It means the two models have the same classification results.

Third, according to the KFCF of the two models, we can find the keywords with higher coefficients in the KFCF of original 3D FOA are regularization and Basic Law, and keywords with higher coefficents in the KFCF of modified 3D FOA are APEC, regularization and Basic Law.

Conclusion

In the beginning, the study used the text mining method to get the keywords idf data of the interviewees' responses. The study utilized the wavelet transformation and attempted to use the 3D FOA model for classification to explore the importance of keywords.

The study concluded that the modified 3D FOA classification had the better performance with regard to the lower RMSE in the iterative optimization process.

Accodring to the KFCF of modified 3D FOA model, the study found the keywords with higher coefficients were APEC, regularization and Basic Law. We can conclude that the keypoints for the interviewees regarding Macau's participation in international organizations include the Basic Law's regularization and Macau's participation in international economic organization.

Acknowledgements

The study was supported by the Macau Foundation Research Fund and Macau University of Science and Technology Faculty

Research Grant.

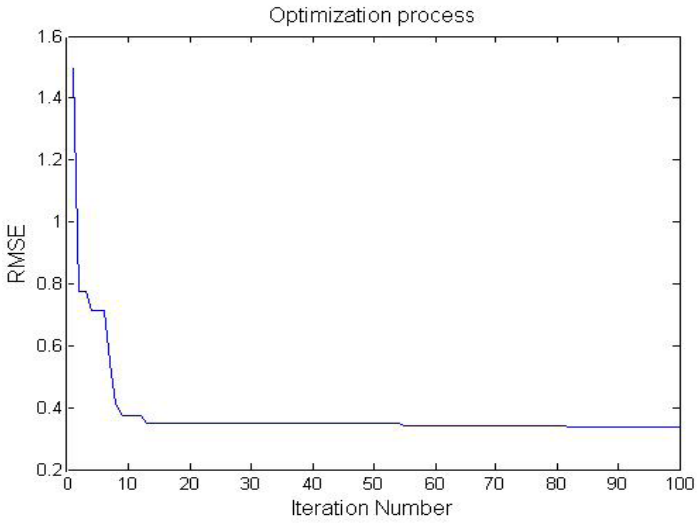

Figure 4. Iterative optimization process of Modified 3D FOA

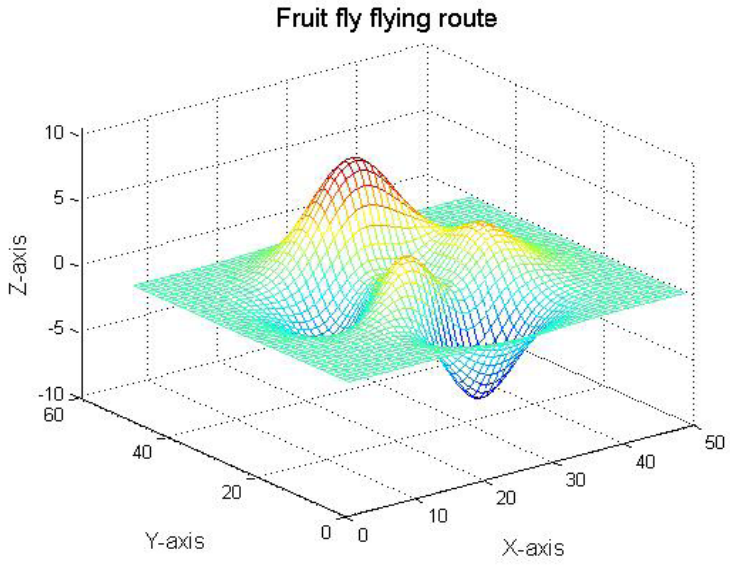

Figure 5. The Searching Route for Fruit Flies of Modified 3D FOA

References

[1] Wolff, S. (2007). Paradiplomacy: scope, opportunities and challenges, *The Bologna Center Journal of International Affairs* 10, 141–150.

[2] Matias, J. C., Macao. (2014). China and Portuguese Speaking Countries China's Macao Transformed: Challenge and Development in the 21st Century, China's Macao Transformed: Challenge and Development in the 21st Century, 2014.

[3] Clement, T. E. (2008). A Thing not Beginning and not Ending: Using Digital Tools to Distant-Read Gertrude Stein's The Making of Americans, Literary and Linguistic Computing 23 (2008), 361–380.

[4] Trappey, A. J. C., Trappey, C.V. (2007). An R & D knowledge management method for patent document summarization, *InduC.V. strial Management Data Systems* 108 (2007), 245–257.

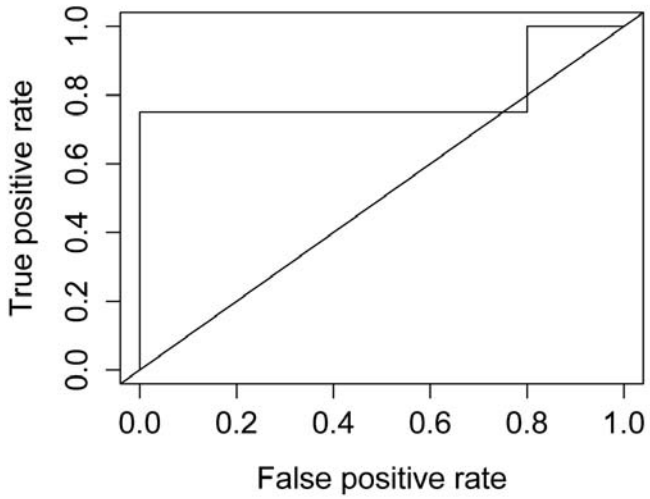

Figure 6. ROC Curve of modified 3D FOA Classification

[5] Schmidt, M. (2010). Quantification of Transcripts from Depth Interviews, Open-Ended Responses and Focus Groups, *International Journal of Market Research* 52, 483–509.

[6] Esuli, A., Sebastiani, F. (2010). Machines that Learn How to Code Open-ended Survey Data, *International Journal of Market Research* 52, 775–800.

[7] Lee, T., BradLow, E. (2011). Automated Marketing Research Using Online Customer Reviews, *Journal of Marketing Research* 48, 881–894.

[8] Li, J. (2013). Rwordseg Usage, https://r-forge.r-project.org/R/?groupid = 1054

[9] Feinerer, I., Hornik, K., Meyer, D. (2008). Text Mining Infrastructure in R, *Journal of Statistical Software*, 25.

[10] Swati, K., Kumar, S. (2015). A Comparative Study of Various Data Transformation Techniques in Data Mining, *International Journal of Scientific Engineering and Technology*, 4, 146–148.

[11] Libungan, L. A., Pálsson, S. (2015). ShapeR: An R Package to Study Otolith Shape Variation among Fish Populations, PloS one,10, e0121102.

[12] Nason, G. (2010). Wavelet methods in statistics with R, Springer Science Business Media.

[13] Pan, W.T. (2012). A new fruit fly optimization algorithm: taking the financial distress model as an example, *Knowledge-Based Systems*, 26, 69–74.

[14] Pan, W.T. (2011). Using Fruit Fly Optimization Algorithm Optimized General Regression Neural Network to Construct the Operating Performance of Enterprises Model, *Journal of Taiyuan University of Technology* (Social Sciences Edition) 4 , 002.

[15] Lin, W.Y. (2015). A novel 3D fruit fly optimization algorithm and its applications in economics, *Neural Computing and Applications*, 1–23.

[16] Lin, W.Y., Huang, T. H. (2014). A Hybrid Approach of 3-D Fruit Fly Optimization Algorithm and General Regression Neural Network for Financial Distress Forecasting, *International Journal of Digital Content Technology and its Applications* 8, 1.

[17] Wu, S. H., Shi, Y.D. (2013). Fruit fly optimization algorithm for keywords frequency composite function of macau's gambling industry, *International Journal of Applied Mathematics and Statistics* 50, 94–100.

[18] Li, C., Xu, S., Li, W., Hu, L. (2012). A Novel Modified Fly Optimization Algorithm for Designing the Self- Tuning Proportional Integral Derivative Controller, *Journal of Convergence Information Technology* 7 (2012), 16.

[19] Mousavi, S. M., Alikar, N., AkhavanNiaki, S. T. (2015). An improved fruit fly optimization algorithm to solve the homogeneous fuzzy series–parallel redundancy allocation problem under discount strategies, *Soft Computing*, 1–27.