



A Study on the Dynamic Characteristics of Willingness to Communicate—Student Clustering Analysis Based on EM Algorithm

Liang Gu and Jie Wang(✉)

Maanshan Teacher's College, Maanshan 243000, Anhui, China
wj@massz.edu.cn

Abstract. Based on the complex dynamical systems theory, this study used the EM iterative algorithm in order to study the willingness to communicate of second language learners from a dynamic perspective. The students are assigned to a certain cluster according to the characteristics of their willingness to communicate and the representative categories of their willingness to communicate are classified. Then the cluster analysis is carried out. The results show that the diversity of different groups' willingness to communicate is obvious, showing a dynamic development, and has a high correlation with the level of oral proficiency. The research has certain guiding significance for seeking the countermeasure to promote the willingness to communicate, deepening the research of Oral English teaching and improving the language teaching method.

Keywords: Complicated Dynamical systems theory (CDST) · Characteristics of willingness to communicate · Em Algorithm · cluster analysis(CA)

1 Introduction

So far, studies have been conducted to track the dynamic development and internal interaction patterns of complexity, accuracy and fluency (CAF) in oral English with the help of complex dynamic system theory (CDST)(Yu & Lowie, 2020; Polat & Kim, 2014; Yu Hanjing, 2020; Zheng Yongyan, Liu Feifeng, 2020), The findings enrich the research on L2 oral dynamic development and deepen the understanding of L2 learners' oral competence by language researchers and teachers, which proves the applicability and validity of CDST in the study of research on second language Development [1]. At present, most studies can investigate the dynamic development of language dimensions, as well as the internal and external factors of the fluctuation of oral communication level. However, few studies have explored the reasons which leads to the difference of oral English level or the fluctuation of communication willingness at different stages of the same individual from the perspective of "Dynamic", neglecting the interaction and relationship between the willingness to communicate as a situational variable with these internal and external factors, which make the study lack of comprehensive and systematic.

To determine the important factors that affect the system development is one of the core tasks of CDST research. Both the intrinsic factor (motivation) and extrinsic factor (learning situations) influence the development of the complex system of language by affecting the willingness to communicate. In view of this, based on CDST, this research uses the clustering analysis of EM algorithm, the combination of quantitative research and qualitative research to observe the communication willingness of the research objects from a dynamic perspective, and summarizes the dynamic characteristics of the communication willingness, identify the representative categories of willingness to communicate. It is expected that the teacher can find the corresponding countermeasures to adjust effectively, optimize the oral teaching methods and improve the students' oral English level.

2 Complex Dynamical Systems Theory

2.1 The Introduction, Connotation, Theoretical Basis and Development Trend of CDST

In 1997, complex Dynamical systems theory (de Bot & Larsen-freeman 2011; Larsen-freeman 2015; Larsen-Freeman & Cameron 2008; Ortega & Han 2017) was first introduced into applied linguistics research by Professor Diane Larsen-freeman of the University of Michigan. The theory was originally applied to mathematics, advocating a systematic view of the problem, that is, by understanding the relationship among parts in order to grasp the whole [2, 3].

The introduction of this theory has just started in China. Papers published in foreign language journals include Shen Changhong, Lu Min (2008), Wang Tao (2010, 2011), Li Lanxia (2011), Zheng Yongyan (2011), Wei Xiaobao (2012), Dai Yuncai and Wang Tongshun (2012) [4]. Its theoretical basis is the self-organization theory of the 1970s, Prigogine's synergetics and M. Eigen's hypercycle theory [5]. That is, complex phenomena are not static and isolated states, but interactive and dynamic processes in the perspective of "Complex Dynamical systems theory". The system theory has ten characteristics, there are: dynamic, complex, nonlinear, chaotic, unpredictable, sensitive at the initial state, developed, self-organized, sensitive feedback and self-adaptation [6].

In recent years, the theory shows a new trend of interdisciplinary integration and as a latest theory in the field of second language acquisition (SLA) research, which has attracted much attention from both the international and domestic applied linguistics [7].

2.2 Dynamic Changes in the Characteristics of the Willingness to Communicate Based on Complex Dynamical Systems Theory

Different from previous studies, this study uses qualitative and quantitative methods (such as Han & Hiver, 2018) based on the CDST perspective to simplify the research subjects, by long-term tracking and in-depth data mining of individual learners' communication willingness, this paper studies the importance of the communication willingness as a situation variable in second language learning from a "Dynamic" perspective, which

can present the state of learners' willingness to communicate and the process of change and development in the real language learning environment. The results show that learners' willingness to communicate varies, constantly changes and develops in different ways under the influence of various factors.

The study has some enlightening significance on the field of tracing the development mode further and influencing factors of communication willingness, the influence of communication willingness mode on oral English learning, and instructs teachers to seek corresponding strategies for effective adjustment in order to promote benign development.

3 Study on Clustering Analysis for the Dynamic Feature of Communication Willingness Based on EM Algorithm

3.1 EM Algorithm and Process

Em (Expectation Maximization) algorithm is an excellent algorithm to solve the problem of data incompleteness. The most common problem of data incompleteness is that the data contains hidden variables that can not be directly observed [8].

In this study, when we use the EM algorithm to cluster the acquired data, we do not assign the objects (refer to students in this paper) to a certain cluster, instead, it allocates objects according to the probability of the affiliation between objects and clusters [9]. In the data mining, the variables (attributes) of students' communication behavior, manner, habit and preference are set to be observed and tested directly through certain questionnaires and tests, They are accessible data, while the students' willingness to communicate represented by these data is a hidden variables which can not be observed directly and needs to be discovered. Therefore, the whole clustering procedure need to be finished based on the occurrence of the above learners' willingness to communicate.

The EM algorithm is essentially an iterative algorithm to cluster the dynamic features of willingness to communicate. For the observed Data X, it is not known which branch it is from, so it is incomplete data. The complete data should be expressed as (x, y) , where y denotes the label of the branch to which X belongs and takes a range of value $y \in \{1, \dots, g\}$. The probability density of the overall data can be expressed as:

$$f(x, y; \theta) = \sum_i^g r_i f_i(x, y; \theta_i)$$

In the above formula, "g" is the number of density branches; r_1, r_2, \dots, r_g is the proportion of the total distribution of each branch point, $\sum_i^g r_i = 1$; " f_i " is the density of the i^{th} branch, " θ_i " is the unknown parameter of the corresponding branch. With the observation data sets $\{x_1, x_2, \dots, x_n\}$, the value is evaluated using the maximum likelihood method.

It starts with the initial solution " θ_0 ", for $t = 0, 1, 2, \dots$ by repeating iteratively to get the $\theta_1, \dots, \theta_t$.

In each iteration, the expected value of the Logarithmic likelihood function for the complete data is calculated with the observed data and the current solution " θ_t " given. Then, a new parameter " θ_{t+1} " is chosen to maximize the expected value of the Log-likelihood function. And then we iterate back and forth, and we get " θ_t ", until we get convergence. As shown in the Fig. 1.

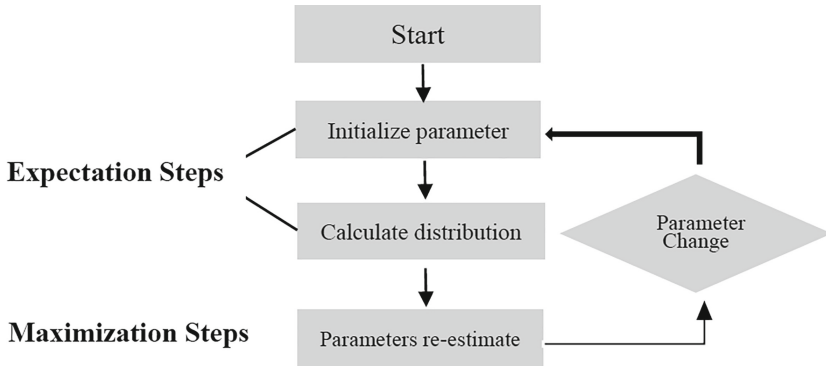


Fig. 1. An iterative flowchart (Self-drawing)

3.2 Test and Classify the Characteristics of Communication Willingness

Data from 1. The communication willingness questionnaire (McCroskey, 1992), $\alpha = 952$, consisted of 24 questions. Participants were asked to rate their possibility of interpersonal communication freely with others in Chinese/English on a scale ranging from 0 to 100, 0 means “Never”, 100 means “Always”; 2. Self-assessment of communicative competence by McCroskey (1988), $\alpha = 867$, with 12 items, was assessed on a scale ranging from 0(no communication ability at all) to 100(having complete communication ability), rate your ability to communicate in Chinese/English on a given occasion. The social situations include public speech, delivering on a large conference, speaking in a group; strangers, acquaintances and friends are all involved [10]. Then, according to the degree of integration score, the willingness to communicate was divided into three types: High (H) level, Medium (M) level and Low (L) level.

4 Conclusion

Based on the CDST, the research used the EM algorithm to make a cluster analysis (CA) on the communicative willingness of language learning groups, and got a good result. There are obvious characteristics in the choice of communicative willingness and self-assessment of communicative competence of different types of students in different situations when they use English or Chinese, which are basically consistent with the reality. Here are what the study found:

- i. The communicative willingness of the learning group is influenced by some factors, which shows the characteristics of diversity, and has a certain correlation with the level of oral English proficiency;
- ii. Based on the data mining of students’ communicative behavior, manner, habit and preference, the representative categories of communicative willingness can be classified by clustering analysis with EM algorithm;
- iii. At different stages of the same individual, the willingness to communicate changes and shows a dynamic development.

The study of this paper made a preliminary attempt to the study of the dynamic characteristics and cluster analysis of communicative willingness, which provided a reference for tracking the dynamic development trend and influencing factors of learners' willingness to communicate in the next stage and exploring the law of development of learners' willingness to communicate comprehensively.

Acknowledgements. The exploration and practice of foreign language ideological and political education in higher vocational colleges under the background of new liberal arts "taking Foreign reception module teaching as an example" (2022jyxm1553); Major research project in natural science of Anhui Province: Research on edge computing framework and optimization mechanism for industrial internet (2022AH040346); Education and Teaching Research Project of Anhui Province, China "A study on the development of intercultural foreign language teaching competence of applied undergraduate teachers based on Community of Practice" (2021jyxm1006).

References

1. Yu Hanjing, Liu Tianqi. A study of typical dynamic features of learners' oral motivation based on the retrospective paradigm approach [J]. *Foreign Languages and foreign language teaching*, 2021(4): 68-79
2. Mason, M. (2008). What is complexity theory and what are its implications for educational change? *Educational Philosophy and Theory*, 40(1): 35-49.
3. Larsen-Freeman, D.(1997). Chaos/ Complexity science and second language aquisition. *Applied Linguistics*, 18(2): 141-165.
4. Xu Xiyang, Wu Yongyi. Complex Dynamical systems theory: Reflections on the study of second language acquisition [J]. *Language teaching and research*, 2015(2): 1-7.
5. Liu Wenyu, Cheng Yonghong. A Review of Applied Linguistics Research from the Perspective of Complex Dynamic Systems Theory [J]. *Contemporary Foreign Language Research*, 2013(8): 31-34+77.
6. Liu Zhencong, Diao Huiying, Li Ting. Revelation of complex Dynamical systems theory for education reform [J]. *Think Tank Times*, 2019(10): 288-291.
7. Zheng Yongyan. Review and Comparison of Theoretical Research on Complex Dynamic Systems in the past Decade [J]. *Research on Second Language Learning* 2020(01): 84-98+119.
8. Dempster A P, Laird N M, Rubin D B. Maximum Likelihood from Incomplete Data via the EM Algorithm [J]. *Journal of the Royal Statistical Society, Series B*, 1977, (39):138.
9. Zhang Chi, Chen Gang, Wang Minjuan. Student clustering analysis using EM algorithm in Mobile Learning [J]. *Technology applications*, 2009(5): 68-71.
10. Wang Miao, A study of willingness to communicate in second language classrooms [J]. *Journal of Hunan Vocational College of Post and telecommunications*, 2015(12): 128-133.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

