# A Study on Peer-Assessment Based on Collaborative Out-of-Class Learning

Meixiu Lu, Manzhen Yang, and Xia Li

Abstract-Peer assessment can help address the problem of imbalance between teachers and students in China's higher education environment, and it can also effectively promote the development of learners' various abilities. However, previous studies indicate that students generally ascribe more value to teacher assessment than to peer assessment. Accordingly, methods to improve the efficiency of peer assessment demand more attention. This study explored the validity of peer-assessment and the acceptance of peer assessment via case collaborative out-of-class learning in a "natural language processing" course. Experiments showed a statistically significant correlation, but no significant difference, for peer-assessment between-groups and the teacher assessment. Peer-assessment can be used as a reference for the teacher assessment. On the contrary, there was a statistically significant difference but no significant correlation for the within-group assessment and the teacher assessment. Additionally, students reported that they found the peer-assessment based on collaborative out-of-class learning helpful and felt they benefited from peer assessment during teaching practice.

*Index Terms*—Collaborative out-of-class learning, natural language processing, peer-assessment, engineering education.

### I. INTRODUCTION

Cooperative out-of-class learning, first proposed in the 1970s by famous American educator David Koonts, has developed into one of the most respected teaching theories [1]. At the end of the 1980s, scholars pointed out that communication, consultation and cooperation were important factors in cultivating learners' autonomous learning ability. Vygotsky's social and cultural theory shed light on how learning is shifted from social sharing activities to individual internalization processes [2]. In the social context of "interaction, mutual support, [and] mutual coordination" composed of "teachers, peers, and collectives and other elements" [3], learners realized the internalization of autonomous learning ability through mutual dependence, mutual cooperation and friendly communication. Vygotsky's Zone of Proximal Development (ZDP) provides the theoretical basis for peer assessment. ZDP refers to the distance between the individual development level needed to solve a problem independently and the potential development level which is needed to solve a problem under the guidance of an adult or collaboration with a stronger peer [4]. Closing this distance means learning occurred. Moreover, the form of mutual learning during the extracurricular cooperative can provide an environment and conditions for learning.

Peer assessment has become increasingly popular in education fields and is the key issue in formative assessments. Peer feedback can be timelier and more individualized than teacher assessment. For example, peer assessment requires learners to provide teachers with a more accurate description of an individual's performance in group work [5]. and evaluates peers' work or academic performance via the same learning environment [6]. Topping (1998) describes peer assessment as the learners evaluating how the peers' learning functions or learning results in the same level, based on the difficulty, value, utility, quality, and other aspects of products [7]. Through learning content in the same course, and having common learning objectives and learning tasks, the peers comprise the same learning environment or could be said to have the same level. With the exchange, support, mutual learning tasks and learning scenarios, the learners can more deeply understand peers' learning process. Peer assessment can solve problems relating to the shortage of teachers, the difficulty of teaching a diversity of subject areas, and inadequate teacher feedback. Related work may consist of conducting peer assessment, including writing, oral presentations, portfolios, test performance, or other skilled behaviors.

However, whether peer assessment can replace teacher assessment, and the degree of subjective acceptance of the peer assessment, has not been studied during the process of peer assessment implementation. Indeed, there remains a lack of empirical research regarding the validity of peer assessment in the field of engineering education in higher education in China. This study aimed at exploring the effectiveness of peer assessment and students' acceptance of peer assessment in an extracurricular cooperative learning environment.

#### II. REVIEW OF RELATED LITERATURE

Some related research reviews have already conducted and the focus has been from the theoretical exploration and empirical research to the discipline teaching system construction process. Marjo *et al.* (2010) conducted a study of 26 typical high-level peer reviewed articles published by SSCI search journals over the past 17 years and found that 25 were conducted in a higher education environment [8]. Higher education environments do not have high-risk examinations, and are more suitable for peer-related research.

Manuscript received October 26, 2017; revised May 12, 2018.

Meixiu Lu is with School of Information Science and Technology/School of Cyber Security, Guangdong University of Foreign Studies, Guangzhou, 510420, China (e-mail: lmx1977@163.com).

Manzhen Yang is with School of English and Education, Guangdong University of Foreign Studies, Guangzhou, 510420, China (corresponding author: Manzhen Yang; e-mail:199610576@oamail.gdufs.edu.cn).

Xia Li is with Laboratory of Language Engineering and Computing, Guangdong University of Foreign Studies, Guangzhou, 510420, China (e-mail: 200211025@oamail.gdufs.edu.cn).

Additionally, there were some peer assessment systems applied to teaching practice. However, most of these studies focused on language learning and teaching. For example, Calibrated Peer Review (CPR) is an online platform used to conduct peer assessment, in which students can submit essays online and are guided to rate essays using criteria that instructors have written [9]. Another peer-review system, Scaffolded Writing and Rewriting in the Discipline (SWoRD), also called Peerceptiv, allows student to conduct peer assessment and exchange ideas online [10].

Some studies focused on peer assessment in teaching practice have found that there are statistically significant differences between teacher assessment and peer assessment [11], [12]. Peer assessment scores were found to be higher than teacher scores, and there was a statistically significant difference in peer assessment from different individuals using the same product [5], [13].

The study of peer reviews in China is still in the beginning stages. Some scholars have conducted related research, focusing on the influencing factors of peer assessment implementation, the peer assessment process, and peer assessment strategy. Factors of interest, individual differences and execution cost represent three factors which affect the reliability and validity of peer assessment through principal component factor analysis [14]. Ling Zeng et al. (2017) attempted to implement teaching practice with peer assessment using the Moodle3.0 platform [15]. Zhiqiang Ma et al. (2016) explored the relationship between the final peer review text, review types and the availability of judgments, using content analysis and regression [16]. Hongxing Jiang et al. (2016) proposed the basic process and specific strategies of assessment models regarding electronic works based on the network environment [17]. Furthermore, one study (Yan Liu et al., 2016) found that peer assessment had a positive effect on developing students' critical thinking skills [18].

However, little attention has been placed on the validity of peer assessment based on extracurricular cooperative learning and student attitudes on peer assessment in China. Under existing conditions, each teacher may have up to 100 or more students; with less time and energy, teachers cannot efficiently conduct assessments of each student's work. A learning mode based on extracurricular cooperation, which is more suited for peer assessment in the process of engineering teaching, is explored.

The key issues in effectively implementing peer assessment include: making full use of the advantages of peer assessment, giving full play to its role in promoting learning, investigating the subjective attitude of students and proposing targeted solutions. Based on the above dynamics, this study carried out a peer assessment empirical research study based on an extracurricular cooperative learning environment with a "natural language processing" course. The specific research questions were as follows:

- 1) What is the validity of peer assessment based on extracurricular cooperative learning?
- 2) What are students' perceptions in relation to peer assessment during extracurricular cooperative learning?

## III. METHODOLOGY

# A. Participants and Procedure

Participants include 86 undergraduate students from a university in China with computer science and technology, network engineering, and software engineering backgrounds during extracurricular cooperative learning environment. This study uses a "natural language processing" course as an example. The course lasted four months, amounting to a total of 16 weeks at 3 hours per week, totaling 48 hours, between March 1 and July 1, 2017.

On one hand, junior students are more mature in the professional knowledge system structure after nearly two years of learning. Teachers and students constitute a learning and living environment such that the teacher and student are more familiar with each other after two years of learning. On the other hand, there are no natural administrative classes, and participants enrolled in professional elective courses come from different majors suitable for extracurricular cooperative learning courses. Furthermore, the "natural language processing" course represents interdisciplinary science, and t he learning process is more complicated and suitable for cooperative learning within groups during extra-curricular learning.

The specific sample of this study was obtained as follows. First, during the basic knowledge learning stage, students are required to study the course, in the form of extracurricular cooperation groups, in the first week. The group strategy involved a free combination of students; each group of 3-4 people determined the team leader. Each class further established a Webchat group (a chat tool for instant communication in China) to provide a convenient platform for immediate exchange. The students would complete homework after the study in the form of an extracurricular cooperation group. The teacher selected good and bad examples and asked the students to perform the first round of peer assessment experiments in an experimental class. Teachers explained and compared the work, and gradually helped the students transition from teacher assessment to peer assessment, for the purpose of making students familiar with the peer assessment criteria and processes.

Next, at the course design stage, the students would choose the curriculum design topic in the midterm under the teachers' guidance. Students conducted topic selection, data collection, system design, and implementation in the form of an extracurricular cooperation group learning. The peer assessment, in fixed time and place, can improve the assessment quality. The data are derived from the second round of peer assessment. During the last class, the student would conduct peer assessment with the help of a network teaching platform. The data comes from the implementation of peer assessment in the end of the course assessment data: valid samples of two different classes (N1 = 42 and N2 = 44).

The first part of the data consists of peer assessment scores and teacher assessment scores; the second part consists of questionnaire data. Peer assessment score come from z total of 26 courses designer products in two classes. During the peer assessment stage, each group of students makes a presentation, then gets the scores from all the students in the class (including within-groups and between-groups scores). The questionnaire data comes from the peer assessment, the students' questionnaire results. This analysis was conducted using SPSS 20.0.

# IV. DATA ANALYSIS

### A. Teacher Assessment vs. Peer Assessment

This study explores the validity of peer assessment using t-tests (Table I). Statistical data show no statistically significant difference between peer assessment and teacher assessment (Sig. = 0.00 < .05). Peer assessment can be used as an effective reference for teacher evaluation because there was a fairly strong correlation between teacher assessment and peer assessment. The correlations were r = .548 and r = .688.

The results reveal that the standard deviation of teachers 'scores was significantly higher than that of the peer reviews. Fig. 1 shows that the teachers' scores have a curved distribution, and the scores of peer scores are smooth and linear. The results further indicate a large standard deviation for teachers. The score data are more discrete and distinguishable, indicating that the teacher score reflects the students' academic level more than peer score. However, the standard deviation of peer evaluation is small, and the score data show an aggregation trend. The minimum score of 84, the maximum 93 points, demonstrates that the peer data had a steady trend, and the student level was difficult to distinguish. Therefore, peer assessment may not be a good response to students' actual academic level. This may be because peer assessments scores are affected by other factors. Some related research found that learners will be evaluated more carefully and sincerely if the real name is provided [6], but "privatization" may affect the validity of the results. To further explore this, we conducted additional statistical analyses within-groups and between-groups.

		Descriptive statistics		pearson correlation		
		Mean	SD	Р.	Sig.	t
class1(	peer	87.51	1.713	0.540	0.0001	0.61
N=42)	teacher	87.19	4.038	0.548		0.61
class2( N=44)	peer	87.4	1.456	0.600	0	1.2
	teacher	87.93	3.884	0.688		-1.2

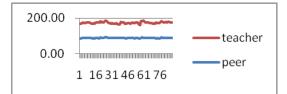


Fig. 1. Comparison between teacher assessment and peer assessment.

# B. Between Groups vs. within Groups

This study uses t-tests to further study differences between-groups and within-groups (Table II). The data show that the standard deviation between groups is less than that within-groups scores. The within-group scores are obtained from the members of the group. The larger the standard deviation, the more discretized the data, the more the members of the group understand the quality of their group's work, and the higher distinction the score has. The scores of the between-groups come from the scoring of all members of the group within the class; the smaller the standard deviation, the more stable the data, and the lower distinction the score has. Which scoring method is more effective, and what is the impact of the factors? To further explore the factors of between groups assessment, this study conducts related research in the form of an open-ended questionnaire. The results demonstrate that other people's understanding of the system and student relationship are the factors most affecting the results. The same study is conducted within groups, finding that most students have a strong collective concept: "a group is a collective honor; should not distinguish between each other," "the feelings between peers," among other factors. To further explore the phenomenon behind the data, this study conducts an additional analysis, finding that 37.2% of the sample within-groups scores were the same, indicating that peer assessment within-groups may not reflect individual contributions. Because the course design requires teamwork and a division of labor to complete different functions, each member's contribution cannot be easily quantified.

TABLE II: PEER ASSESSMENT WITHIN GROUPS VS. BETWEEN GROUPS

		Descriptive statistics		-	arson elation	T-test		
		Mean	SD	P.	Sig.	t	df	Sig.
class1( N=42)	within groups	89.98	3.81	- 0.0	0.69	4.26	41	0
	between groups	87.28	1.79	6				
class2( N=44)	within groups	90.23	4.2	0.3	0.02	5.25	43	0
	between groups	87.13	1.45	6				

 TABLE III: TEACHER ASSESSMENT VS. PEER ASSESSMENT BETWEEN

 GROUPS VS. PEER ASSESSMENT WITHIN GROUPS

	pearson correlatio Descriptive statistics n			T-test				
	Mean	SD	N	P.	Sig	t	df	Sig.
teacher assessment	88	3.954	<u>.</u>	0.3	0.0		- 0	
within groups	90	3.995	86	65	01	-5.3	58	0
teacher assessment	88	3.954	86	0.5 48	0	1.02	85	0.3
between groups	87	1.447						

Statistical data analysis illustrates statistically significant differences between-groups and within-groups, but no

significant difference among teacher assessments and between-groups after using a paired-samples t-test (Table III). Peer assessment between-groups are more objective. Anonymous assessment is adopted in this study. The within-groups assessment is affected by subjective factors. Statistical analysis demonstrates a statistically significant difference between teacher assessment and within-groups assessment. To further study the acceptance of the peer assessment, this study conducted additional data analysis.

## C. Student 's Attitudes against Peer Assessment

Students' acceptance of the assessment method affects the teacher's classroom assessment. Cheng & Sun (2015) found that teacher assessment was significantly correlated with the evaluation type, while the degree of study, learning habits, and normal performance of students were also factors [19]. Moreover, the students who participate in peer assessment have a stronger learning motivation and interest in both their learning tasks and peer work evaluation [20], [21]. To further explore the acceptance of students in the context of cooperative out-of-class learning, this study used a comparable questionnaire (Table IV) based on work from Lisa-Marie Collimore [19] and the acceptance model of the user's acceptance of a particular technology in a technical field [22].

This study conducts a questionnaire-based statistical analysis to explore student perceptions of peer assessment during extracurricular cooperative learning. Results illustrate that 80% of the students in both classes had positive attitudes on peer assessment, while only a few students hold a negative attitude. Questions 1-4 examine whether peer assessment contributed to learning; questions 5-8 examine students' understanding of peer assessment; and questions 9-12 examine students' recognition of extracurricular cooperative learning. This study summarizes the 1018 options for 12 questions (Fig. 2). Questionnaire analyses (means and standard deviations) for the five-point items are presented in Table IV.

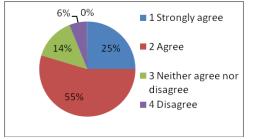


Fig. 2. Distribution of student feedback.

TABLE IV: MEAN AND STANDARD DEVIATIONS FROM STUDENTS' OUESTIONNAIRE

Num	Item	Mean
1	I think peer assessment is more effective than teacher assessment.	2.61
2	I think peer assessment is good for learning.	2.02
3	I think that peer assessments of different works are more helpful in promoting learning.	1.82
4	I think this activity is a process for my submission before I participate peer assessment.	2.32

5	I think this activity is from a different point of view to understand the same concept before I participate peer assessment.	2.14
6	I think this activity is a process for my submission after I participate peer assessment.	2.55
7	I think this activity is from a different point of view to understand the same concept before after I participate peer assessment.	2.05
8	I think extracurricular cooperative learning helps improve my ability to learn autonomously.	1.66
9	I think extracurricular cooperative learning has expanded my thinking ability.	1.65
10	I think that extracurricular collaborative learning can help improve the effectiveness of peer assessment.	1.80
11	I think extracurricular cooperative learning has improved my enthusiasm for learning natural language courses.	1.73
12	I think it helps to modify and resubmit my work after peer assessment.	1.85

#### V. DISCUSSION

Statistical analysis indicates no statistically significant difference between teacher assessment and peer assessment between-groups; however, there is a statistically significant difference between teacher assessment and peer assessment within-groups. Therefore, this study addresses the first research problem; peer assessment based on extracurricular cooperative learning can be used as an effective reference for teacher assessment. Nevertheless, there are many factors during the peer assessment process, such as the differences between-groups and within-groups, and the calculation of the personalized contribution in teamwork.

Considering the second research question: this study designed an open-ended questionnaire with the query "Which way is more appropriate for peer assessment?", and performed further statistical analysis. The results here shows that 87.06 % o f the students believe that anonymous peer assessment is more effective, which is consistent with the results of the previous analysis demonstrating no significant difference between teacher assessment and peer assessment between groups; only 11% of the students believe using one's real name was more appropriate. To explore these reasons more in-depth, this study employed other open-ended questions, e.g., "What is the key factor that will affect the implementation of peer assessment?" and "What has been learned from peer assessment process?" Results showed that the factors affecting peer assessment were: assessment platform, classmate friendship, the degree of understanding of the products, whether the peer assessment was anonymous, and other factors. Most students believed that the following would be learned during the process of peer assessment understanding their own deficiencies, getting а comprehensive view of the problem, having a proliferation of thinking, learning from each other, learning teamwork, and developing communication skills. The implications from our study therefore relate not only to promoting students'

learning based on extracurricular cooperative learning, but also to producing a process accompanied by the learning process.

## VI. CONCLUSION

This study explores the validity of peer assessment and the acceptance of peer assessment in collaborative out-of-class learning. Peer assessment implementation is affected by a number of factors. Although teacher assessment is considered the most effective assessment method, teachers often do not have enough time and energy to implement efficient assessment in the teaching process, given their dual pressures of teaching and researching. Peer assessment, an effective assessment method in the implementation of formative assessment, has demonstrated its value. The results of this study show that teachers and students possess positive attitudes about, and accept, the peer assessment process.

The statistical analysis indicates no statistically significant differences between teacher assessment and peer assessment between groups, an effective reference for teacher assessment. Furthermore, modern network technology will be used to more efficiently implement peer assessment between groups. Considering students' attitudes, most students accept peer assessment, but there are some problems in the specific implementation process. The teacher should take measures, such as using anonymous peer assessment and assessing during a specific time and place, to avoid negative effects and effectively conduct peer assessment.

Peer assessment, to a certain extent, can effectively promote the cultivation of students' self-regulated learning abilities and can reduce the teachers' burdens. At the same time, peer assessment provides students with the opportunity to learn from one other and promotes "assessment for learning." Peer assessment focuses on learning process, not the results, closing the gap between the learning goal and the existing level during the process of learning. Through the described processes, students can achieve self-regulated learning and have gradually reduced dependence on teachers [23].

Limitations of this study relate to the small number of students who participated in the study, which prevented us from generalizing our conclusions. Another limitation of this study includes the course selection, which includes intercultural courses, and cannot be generalized to a larger population. The development of new technologies will bring new opportunities for implementing peer assessment. Whether learning in an online community will be helpful in cultivating self-learning ability, and in improving the validity of peer assessment, will be further explored. Follow-up research here will focus on the grouping strategy of different learning styles in online communities.

To effectively conduct peer assessment, this study draws some future research directions. First, high quality peer assessment requires the support of the network environment in providing a safe and free environment for peer assessment. Second, peer assessment requires artificial intelligence technology to conduct more in-depth mining with students' learning data and to complete dynamic assessment, timely feedback visualization to the teacher, thereby enabling the teacher to adjust teaching decisions.

## ACKNOWLEDGMENT

**The funding of this research was supported by** National Science Foundation of China (61402119,61572145);

Teaching Quality and Teaching Reform Project in Guangdong Province, China (236:No.218); Teaching Quality and Teaching Reform Project in Guangdong Province, China(175:No.257); Guangdong Province Database series course teaching team(233:No.20); Course Construction Project of Practical Talents Training in Guangdong Province(Natural Language Processing); Guangdong Provincial Department of Education key platform and scientific research project characteristics of innovative projects(2016KTSCX036)

#### REFERENCES

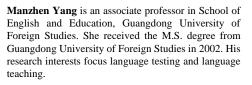
- X. U. Jin-Fen, "The effect of collaborative out-of-class learning on the development of students' autonomous English learning ability[J]," *Journal of Pla University of Foreign Languages*, vol. 5, pp. 39-43, Oct 2013.
- [2] G.-S. Xiao *et al.*, "A study on the relationship between college students' social support, class attribution and English autonomous learning," *Foreign Language World*, vol. 4, pp. 2-11, December 2011.
- [3] F.-H. Zhang *et al.*, "Charting the course of the theoretical exploration of peer reviews in writing," *Journal of Northeast Normal University(Social Science)*, vol. 4, pp. 226-229, August 2015.
- [4] W. Cheng and M. Warren, "Making a difference: Using peers to assess individual students' contributions to a group project[J]," *Teaching in Higher Education*, vol. 5, pp. 243-255, August 2010.
- [5] C. C. Chang *et al.*, "Reliability and validity of web-based portfolio peer assessment: A case study for a senior high school's students taking computer course[J]," *Computers & Education*, vol. 57, issue 1, pp. 1306-1316, August 2011.
- [6] F. Y. Yu and C. P. Wu, "Different identity revelation modes in an online peer-assessment learning environment: Effects on perceptions toward assessors, classroom climate and learning activities[J]," *Computers & Education*, vol. 57, issue 3, pp. 2167-2177, November 2011.
- [7] K. Topping, "Peer assessment between students in colleges and universities[J]," *Review of Educational Research*, vol. 68, pp. 249-276, Fall 1998.
- [8] M. V. Zundert, D. Sluijsmans, and J. V. Merri öboer, "Effective peer assessment processes: Research findings and future directions[J]," *Learning & Instruction*, vol. 20, pp. 270-279, August 2010.
- [9] L. Likkel, "Calibrated peer review essays increase student confidence in assessing their own writing[J]," *Journal of College Science Teaching*, vol. 41, pp. 42-47, January/February 2012.
- [10] K. Cho and C. D. Schunn, "Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system[J]," *Computers* & *Education*, vol. 48, pp. 409-426, April 2007.
- [11] T. Hovardas, O. E. Tsivitanidou, and Z. C. Zacharia, "Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students[J]," *Computers & Education*, vol. 71, pp. 133-152, February 2014.
- [12] C. C. Tsai, S. S. J. Lin, and S. M. Yuan, "Developing science activities through a networked peer assessment system[J]," *Computers & Education*, vol. 38, pp. 241-252, January–April 2002.
- [13] C. C. Chang, K. H. Tseng, and S. J. Lou, "A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a web-based portfolio assessment environment for high school students[J]," *Computers & Education*, vol. 58, pp. 303-320, January 2012.
- [14] B.-Y. Yin *et al.*, "Analysis on the influencing factors of correspondence electronic works[J]," *E-education Research*, vol. 12, pp. 58-62, Dec. 2012.
- [15] L. Zeng *et al.*, "A review of the application of complementary evaluation module in moodle 3.0[J]," *Software guide*, vol. 16, pp. 195-198, May 2017.
- [16] Z.-Q. Ma *et al.*, "An analysis of the types and effects of feedback comments in network compilers[J]," *E-education Research*, vol. 1, pp. 66-71, January 2016.

- [17] H.-X. Jiang and M. Qiu, "Research on the mode of complementary evaluation of electronic works in network environment[J]," *Computer Education*, vol. 258, pp. 81-84, June 2016.
- [18] Y. Liu *et al.*, "Using peer teaching method to train students' critical thinking ability[J]," *Experimental Technology and Management*, vol. 33, pp. 155-158, Feb. 2016.
- [19] L. Cheng and Y. Sun, "Teachers' grading decision making: Multiple influencing factors and methods[J]," *Language Assessment Quarterly*, vol. 12, pp. 213-233, May 2015.
- [20] P. Humphreys, K. Greenan, and H. Mcilveen, "Developing work-based transferable skills in a university environment[J]," *Journal of European Industrial Training*, vol. 21, pp.63-69, 1997.
- [21] K. Macpherson, "The development of critical thinking skills in undergraduate supervisory management units: Efficacy of student peer assessment[J]," Assessment & Evaluation in Higher Education, vol. 24, pp.273-284, Sep. 1999.
- [22] X.-Q. Gu and Y.-K. Zhu, "Research on the acceptance of technical products by educational users: Differences between teachers and students and their mutual influence[J]," *E-education Research*, vol. 8, pp. 21-27, Oct 2012.
- [23] W.-X. Zhang and W. Yong, "Assessment for learning: Constructing an automated writing evaluation system based on writing tasks," *Modern Education Technology*, vol. 27, pp. 99-104, March 2017.





Meixiu Lu is currently a PhD candidate, a lecturer in School of Information Science and Technology/ School of Cyber Security, Guangdong University of Foreign Studies. She received the M.S. degree from Guangzhou University in 2003. Her research interests are in areas of natural language processing and teaching technology and assessment.





Xia Li is a professor in Laboratory of Language Engineering and Computing, Guangdong University of Foreign Studies. She received the doctor degree and B.S. degree in 2002, 2017. Her research interests are in areas of natural language processing and data mining.