Exploring the Relationship between Participant Role and Collaborative Quality in Online Collaborative Discussions

https://doi.org/10.3991/ijet.v18i11.38699

Huichao Mi[1], Qiaorong Zhang[1], Yafeng Zheng[2](✉)
[1]College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou, China
[2]A Center for Educational Science and Technology, Institute of Advanced Studies in Humanities and Social Sciences, Beijing Normal University, Zhuhai, Guangdong Province, China
zlzyf@126.com

Abstract—The exploration of the role concept has become an important perspective for analyzing and promoting computer-supported collaborative learning (CSCL). Understanding the relationships between individual participation roles and collaborative performance is of great significance to the research of collaborative learning theory, pedagogy and technology. However, few empirical studies investigated the individual participation roles in collaborative discussions and the impact of participation role configuration on group performance. Based on the interactive content of learners in collaborative discussions, this research uses machine learning methods to automatically identify learners' participating roles. Through cluster analysis, five different roles are identified: leader, problem solver, coordinator, marginal learners and learners with difficulties. Furthermore, this research explores the relationships between individual participation roles and group collaboration quality. The results show that groups with different collaboration performances have different role compositions, and the roles of leader, problem solver and coordinator have significant positive effects on collaboration performance. Learners with difficulties have a negative impact on collaboration performance. Combining the research results with the discussion content of the learners, this research conducted an in-depth discussion and analysis of the characteristics of each role, and proposed implications for teaching guidance and researchers.

Keywords—online collaborative discussion, role analysis, emerging role, clustering analysis

1 Introduction

Collaborative discussion is regarded as an ideal way of collaborative learning. Through discussion, students can learn together, reach consensus, and solve problems [1]. Online discussion is an excellent activity to build knowledge, because the combination of explaining, elaborating and defending one's position to others enables learners to integrate and elaborate knowledge in a way that helps them learn at a higher level [2]. In collaborative discussion, individual participation roles has become an important factor affecting the quality of collaborative learning.

In the past ten years, the concept of role has become one of the important perspectives to analyze and promote the development of CSCL. Research on roles in the collaborative process can provide important help for learners to self-regulate [3], monitor and evaluate the learning process [4], and explore the mechanism of collaborative interaction [5].There are two perspectives on roles in the CSCL literature: emerging roles and scripted roles. While there have been some research efforts investigating script roles in collaborative learning [6, 7], studies on emerging roles have received less attention [4], such as automatic detection of emergent roles, the relationship between emergent roles, collaboration patterns and group performance. Emerging roles are the roles that students form spontaneously in collaborative learning [6]. This view of emerging roles emphasizes the construction and self-regulation of learners in collaborative activities. Analyzing the emerging roles can help us better understand the cooperation mode of members in collaborative discussions and the relationship between the cooperation mode and the collaborative quality [6] [8, 9].

Based on these considerations, we used machine learning and statistical analysis methods to automatically identify the emergent participant roles in online collaborative discussions, analyze the role composition of collaborative groups with different collaboration performance, and explore the relationship between the individual emerging roles, group role composition and collaborative learning effects in online collaborative discussions.

2 Related work

2.1 Participatory roles in collaborative learning

In recent years, the concept of role has become an important concept to promote and evaluate computer-supported collaborative learning (CSCL). Roles in collaborative learning can be defined as the specified functions or responsibilities that can guide individual behavior and regulate the interaction within the group to a certain extent (Hare, 1994). Assigning specific roles to students can promote students' active internal dependence, personal accountability, and cognitive participation in the process of collaboration, and help them reach a higher stage of knowledge construction. In addition, roles can stimulate members' awareness of the performance of the whole team and the contributions of each member. Roles can promote team cohesion [10], positive interdependence and personal responsibility [11], which are the core supporting factors of collaborative learning arrangements [12, 13]. For example, students appointed as leaders show more active participation, spend more time in on-line discussions, and use various methods to interact with other group members.

There is no generally accepted taxonomy of roles [14]. According to the manner of formation, there are mainly two perspectives on roles in CSCL: the emerging roles perspective and the scripted roles perspective [15]. The perspective of scripted roles focuses on how to facilitate a collaborative learning process by constructing and defining learner roles and activities [16]. This perspective emphasizes the need for teaching support, especially to improve collaborative learning processes and outcomes. Script roles are functional because they specify activities that are considered to be related to collaborative processes and knowledge building, and learners rarely participate spontaneously, such as giving explanations, constructing arguments, or effectively resolving conflicts [15]. However, although assigning learners specific roles can effectively improve the collaborative process, allowing group members to play assigned roles will prevent the potential benefits of role flexibility, which is essential for effective collaborative learning. Emerging roles are the roles that students form spontaneously in collaborative learning. The perspective of emerging roles emphasizes the construction and self-regulation of learners in collaborative activities. Analyzing the emerging roles can help us better understand the individual contributions in the group and the interaction model among group members.

2.2 Emerging roles recognition

Emerging roles are roles adopted spontaneously by group members without external intervention. Research on emerging roles can help us better understand the contributions of individuals in the group and the interaction mode among group members [6] [9]. Emerging role recognition and analysis is important for online collaborative learning, which can provide support for online collaborative process analysis, self-regulation of different roles, and evaluation of individual student performance [17].

In computer-supported collaborative learning research, researchers try to identify the individual's role in collaborative learning. These studies usually use interviews or content analysis to identify individual roles. De Laat used content analysis to identify the different emerging roles of participants, and discussed the complexity of the emerging role development and group awareness of participants in asynchronous e-learning discussions in the context of higher education [9]. The methods based on interviews or content analysis are time-consuming and not suitable for large-scale data processing. Recently, the focus has shifted to using SNA to recognize roles. Capuano used social network analysis (SNA) and content analysis to analyze the interactions and expertise levels of learners in the CSCL process [7]. Marcos et al. proposed a framework based on SNA to analyze an authentic CSCL course and detected four types of roles in the collaborative groups: teachers as mentors, teachers as collaborators, isolated learners, and collaborative learners [18]. Marcos et al. proposed a semi-automatic adaptive role support method, describing roles as a combination of SNA indicators, defining and identifying the roles of teachers and students in CSCL [19]. Fan Ouyang and Yu-Hui Chang used SNA to examine students' social participatory roles in a graduate-level semester-long online course and defined six social participatory roles: leader, starter, influencer, mediator, regular, and peripheral [5].

SNA only focuses on location information based on the role interaction relationship but does not consider semantic information. Therefore, the interpretability of identifying roles becomes fuzzy. For example, some roles in the same position may be oriented to the content of collaborative tasks, such as the role of summarizing tasks, or management oriented, such as the coordinator. Therefore, it is necessary to identify roles by analyzing the behavioral intention of interactive content, rather than just through interaction relationships. Few studies have considered the content of discussion and how the content influences interaction behavior intentions and patterns in the collaborative process [6].This study closed this gap. This study automatically detected learners' behavioral intention according to their interaction contents and recognized each learner's participatory role using cluster analysis.

2.3 Emerging roles and collaborative quality

The emergent participant role of individuals in collaborative learning is crucial to the quality of collaborative learning. Simone used metacognitive regulation and role analysis to analyze the roles in collaborative learning process according to the contribution characteristics of individuals in student-led productive collaborative learning, and explored the differences of individual participation modes in teams with different collaborative performance [3]. The results showed that individuals can flexibly adopt multiple roles in high-performance groups. De Laat researched the complexity of emerging participant roles and awareness in asynchronous networked learning in the context of higher education, and analyzed how different roles came into being and how they affected group dynamics [20]. De Laat's research highlighted the impact of tasks on how students constructed collaborative activities, and revealed that students formed different roles in online collaborative activities.

Although some existing studies have focused on individual roles on collaborative learning, few studies have studied how participant roles affect the achievement of group goals, and it is not clear how individual participation roles in the group interact and combine in the process of collaborative learning. In the online collaborative learning environment, different combinations of roles can produce different group outcomes. Therefore, another goal of this research is to explore how individual-level roles and the overall role composition of the group affect group performance in cooperative interaction.

2.4 Research questions

The research aim is to identify emerging roles that develop spontaneously during the collaborative discussion and explore the relationship between the role composition and the group performance. More specifically, the following derived research questions will be investigated:

RQ1: What roles do learners play spontaneously during collaborative discussions?

RQ2: What are the differences in the role composition of groups with different collaboration qualities?

RQ3: How do different participatory roles impact the performance of collaboration groups?

3 Method

3.1 Research data and context

The context of this study was a compulsory course named "Data Structure" for undergraduate students majoring in educational technology. This 18-week course was de-livered with a blended learning mode. Each week, instructors and students were scheduled to meet for a two-hour face-to-face lesson in a traditional classroom and a two-hour online learning activity of programming in a computer laboratory. Altogether, there were 158 students, of whom 105 were female and 53 were male. Their ages ranged from 18 to 20 (M = 19, SD = 0.78). These participants were randomly divided into 31 groups of four to five learners. During online learning activity, the Moodle platform [21] was used as a collaborative learning tool. Instructors designed problem-solving tasks and students solved each of the problems in their own group discussion space. Collaborative discussion online lasted approximately two hours. We collected the discourse data these learners generated while they solved the tasks. In total, 2,546 posts were initially retrieved.

For the first research question, supervised machine learning models were used to automatically clarify each learner's utterance intention. An unsupervised method was used to cluster the learners according to their discourse behaviors. Each category represented a kind of role. To answer the second and third research questions, we used statistical analysis to analyze the role composition of each group and explored the relationship between role composition and group outcome. The detailed process is described in the following subsections.

3.2 Emerging role automatic detection

Cluster analysis has been proven to be helpful in understanding and discovering interaction patterns related to specific learner roles in cooperative problem solving [22, 23]. In this study, the K-means algorithm is used to cluster learners with similar characteristics to identify emergent roles in the process of collaborative discussion. The key to the clustering algorithm is the selection of sample features. This study takes the interactive intention reflected by learners' interactive content as learners' features for clustering analysis. The process of automatic role recognition mainly includes two stages: interactive behavior intention recognition and clustering.

Because of the lack of appropriate methods and tools to automatically mine learners' interaction intentions, most of the existing studies use qualitative manual coding methods, which are time-intensive and impractical for dealing with massive data. To overcome this methodological challenge, text classification models were constructed to automatically identify the online discourse behavior intentions in collaboration discussions. The process mainly consists of four steps.

In the first step, 800 posts were chosen to develop the training data and ground truth for evaluating the performance of machine learning models. Two hundred posts were first randomly sampled from the 800 posts and analyzed to develop the coding scheme by two senior researchers. The coding scheme is shown in Table 1. Then 200 posts were coded independently by two researchers. To assess the inter-rater reliability of their coding, Cohen's kappa was calculated and reached 0.847, indicating a high level of agreement between the two coders. Disagreements between these two researchers were discussed and resolved. Finally, each of the two researchers coded 200 posts on their own.

Table 1. Code schema for behaviour intention identification

Code	Category Name	Explain	Examples
C1	Statement	Provide or introduce new ideas, opinions, recommendations or plans to identify problems	We transform the pictures into a directed network, and use the adjacency matrix to store, then use the depth-first search traversal and breadth-first search traversal to output two paths.
C2	Negotiation	Support , challenge or oppose the views of others	1) I think so. The advantages of the adjacency matrix are… 2) No, I think if you store a directed map, it can be separated…
C3	Asking questions	Put forward the doubts, and express what is not clear	I agree with you, but why is it stored by the adjacency matrix?
C4	Management	Management and reminders of collaboration and time schedule	At first, we should discuss the advantages and disadvantages of various storage methods. Then, we should determine the reasons for the use of adjacency table.
C5	Share feelings	Expression of greetings, mutual introduction, anger, frustration, shock and difficulty	1) Interesting idea! 2) Forgive me for ignoring. It's too difficult!

In the second step, different features of these posts were extracted. We extracted three kinds of features: linguistic features, linguistic inquiry and word count (LIWC) features, and topic features.

In the third step, four machine learning algorithms were employed on the training data: Naïve Bayes, support vector machines (SVMs), convolutional neural networks (CNNs) and recurrent neural networks (RNNs). A 10-fold cross validation was used to evaluate the performance of each algorithm.

In the fourth step, the model with the best performance was applied to the rest of the unlabeled posts to automatically detect the students' utterance intentions.

Finally, according to the utterance intention of each learner in the collaborative discussion process, the K-means algorithm is used to cluster learners with similar characteristics. Each category is regarded as a role, and the speech behavior characteristics of each role are analyzed, to find the characteristics and rules of group collaboration more clearly and provide a more meaningful basis for guiding the collaborative process.

3.3 Role composition analysis of groups with different collaborative qualities

To clarify the relationship between the role composition of the group members and the group's outcome, two teachers who have many years of experience in teaching the course "Data Structure" are employed to score the performances of the groups. The performance of a group is given a score between 0 and 100. Then, the role composition of group members in each group was counted, and the differences in the role compositions of the discussion groups with their different performances were analyzed.

3.4 Impact of roles on group outcomes

To understand the influence of different roles on the quality of group collaborations, the correlation between the quality of group collaborative learning and different roles was analyzed. Then, the roles significantly related to the quality of collaboration were selected as the independent variable, group performance was taken as the dependent variable, and multiple regression analysis was used to further analyze the predictive ability of these roles.

4 Results

4.1 Emerging role recognition (Clustering analysis)

After testing the number of clusters, five were selected as the most appropriate clustering results. Table 2 shows the clustering results and the description of each clustering feature. Each cluster is considered as a role. According to their characteristics, the five clusters are named "learner with difficulties", "marginal learner", "problem solver", "coordinator" and "leader".

Table 2. Clustering results

Cluster Number	Cluster Name(Role)	Scale	Description
Cluster 1	Learner with difficulties	30	The learners who are able to ask questions actively but have some difficulties in learning.
Cluster 2	Marginal learner	62	The learners who rarely speak or occasionally ask questions.
Cluster 3	Problem solver	39	The learners who put forward views and opinions and actively answer the questions of other group members.
Cluster 4	Coordinator	18	The learners who are able to put forward organization plan, ask questions, ask for opinions of group members, manage and remind group members, and actively express their emotional attitude.
Cluster 5	Leader	9	The learners who are able to put forward opinions, actively communicate and discuss with others, ask questions and have certain coordination and management abilities.

To understand the influence of different roles on the quality of group collaborations, the correlation between the quality of group collaborative learning and different roles was analyzed. Then, the roles significantly related to the quality of collaboration were selected as the independent variable, group performance was taken as the dependent variable, and multiple regression analysis was used to further analyze the predictive ability of these roles.

To describe the characteristics of each role more clearly, the average value and standard deviation of each role on each attribute are calculated, and the results are shown in Table 3.

Table 3. Characteristic of each role

	Learner with Difficulties 30/158		**Marginal Learner 62/158**		**Problem Solver 39/158**		**Coordinator 18/158**		**Leader 9/158**	
Behavior	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Statement	4.7667	2.3879	2.0000	1.2675	6.7692	1.9394	3.8333	2.2816	11.1111	3.4440
Negotiation	2.0333	1.9205	0.8548	0.9725	2.0513	1.6214	1.7222	1.5645	7.2222	4.4096
Asking questions	4.9333	1.3113	1.2258	1.0310	1.0000	0.8584	2.0556	1.3492	5.1111	2.2551
Management	0.7333	0.9072	0.1452	0.3551	0.5641	0.6804	3.2222	1.1144	5.2222	2.0480
Share feelings	0.8000	2.1877	0.2419	0.6449	0.2564	0.6774	0.4444	0.7838	3.0000	2.8284

Role 1: Learner with difficulties

Thirty learners participated in the discussion as this role, accounting for 19% of the total number of learners. The obvious characteristic of these students is that the mean values of "statement"(4.7667) and "ask questions" (4.9333) are high and the mean values of "negotiation" (2.0333), "sharing feelings" (0.8000), and "management" (0.7333) are low.

Role 2: Marginal learner

There are 62 students in this role, with the highest proportion among the five roles. The characteristics of these learners are that they have the lowest mean value of "statement" (2.0000), "negotiation" (0.8548), "management" (0.1452) and "sharing feelings" (0.2419), and a lower value of "asking questions" (1.2258).

Role 3: Problem solver

The third role includes 39 students, who have higher values in "statement" (6.7692) and "negotiation" (2.0513), a lower value in "management" (0.5641), and the lowest values in "asking questions" (1.0000) and "share feelings" (0.2564).

Role 4: Coordinator

The fourth role includes 18 students, whose contribution to "management" (3.2222) is particularly outstanding. Their contribution to "asking questions" (2.0556) and "sharing feelings" (0.4444) was moderate, and their contribution to "statement" (3.8333) and "negotiation" (1.7222) was low.

Role 5: Leader

The fifth role contains nine students, and their behavior values are the highest in five roles, which are 11.1111("statement"), 7.2222 ("negotiation"), 5.1111 ("asking questions"), 5.2222 ("management") and 3.0000 ("sharing feelings").

4.2 The role composition of groups with different performances

According to the results of cluster analysis, we can obtain the number of learners of each role in each group, as shown in Table 4. The last column "grade" represents the out-come of each group.

Table 4. The role composition of each group

Group-ID	Learner with Difficulties	Marginal Learner	Problem Solver	Coordinator	Leader	Grade
Group1	2	0	0	1	1	83
Group2	1	2	0	2	0	80
Group3	2	2	0	1	0	69
Group4	0	2	1	2	0	85
Group5	2	1	1	0	0	73
Group6	0	1	3	1	0	87
……	……	……	……	……	……	……

According to the final results of each collaborative group, we regard the top eight groups with higher scores as high-performance groups, and the eight groups with lower scores as low-performance groups [1]. The distribution of roles between the high-performance and low-performance groups is shown in Table 5.

Table 5. Role distribution of high-performance and low-performance groups

Group	Role	Number	Ratio	Mean	Median	SD	SE
Groups with high performance	Leader	8	20.5%	1.000	0.00	1.60	0.567
	Problem Solver	12	31%	1.50	1.50	1.51	0.535
	Coordinator	7	18%	0.875	1.00	0.641	0.227
	Marginal Learner	8	20.5%	1.000	1.00	1.07	0.378
	Learner with Difficulties	4	10%	0.500	0.00	0.756	0.267
Groups with low performance	Leader	0	0%	0.00	0.00	0.00	0.00
	Problem Solver	5	12%	0.625	0.50	0.744	0.263
	Coordinator	3	8%	0.375	0.00	0.518	0.183
	Marginal Learner	16	40%	2.00	2.00	1.41	0.500
	Learner with Difficulties	16	40%	2.00	2.50	1.77	0.627

It can be seen from Table 5 that in the high-performance group, the students who play the role of "problem solver" are the most represented (31%, mean = 1.5), followed by "leader" (20.5%, mean = 1) and the "marginal learner" (20.5%, mean = 1), then the "coordinator" role (18%, mean = 0.875), and the "learner with difficulties" role is the least (10%, mean = 0.5). This means that in these groups, the members of the group have a high level of knowledge conducted intense discussions, and finally achieved good results. Conversely, in the low-performing group, the "marginal learner" role (40%, mean = 2) and the "learner with difficulties" role (40%, mean = 2) have the highest proportion followed by the roles of "problem solver" (12%, mean = 0.625) and "coordinator" (8%, mean = 0.375), and no member assumes the role of "leader". This shows that many members of these groups have low knowledge levels, and nearly half of the learners did not actively participate in the discussion. Only 20% of the members actively participated in the discussion and committed to problem-solving, so the final results of these groups were low.

An independent samples t-test was carried out to analyze the differences in role composition between the high- and low-performance groups, and the results are shown in Table 6.

Table 6. Difference of role composition between high and low performance groups

Role	Statistic	df	p	Mean Difference	SE Difference
Leader	1.764**	14.0	0.001	1.000	0.567
Problem solver	1.469*	14.0	0.04	0.875	0.596
Coordinator	1.717	14.0	0.849	0.500	0.291
Marginal learner	–1.595	14.0	0.554	–1.000	0.627
Lerner with difficulties	–2.201*	14.0	0.01	–1.50	0.681

Notes: *p < .05, **p < .01.

It can be seen from Table 6 that statistically significant differences between the high-performance groups and the low-performance groups appear in three roles ("leader", "problem solver", and "learner with difficulties"). This indicates that in the high-performance groups, the discussion among group members is more focused on "statement" and "negotiation", while the group members in low-performance groups are more focused on "asking questions" which helps solve problems less.

4.3 Impact of different roles on collaborative quality in online collaborative discussions

Spearman correlation tests were conducted on the five roles and collaboration performance, and the results are shown in Table 7. The correlation between collaborative performance and "leader", "problem solver", "coordinator" and "learner with difficulties" was statistically significant ($p < 0.05$). "Problem solver" ($r = 0.359$, $p < 0.05$), "coordinator" ($r = 0.398$, $p < 0.05$) and "leader" ($r = 0.403$, $p < 0.05$) were significantly positively correlated with performance, while "learner with difficulties" ($r = -0.490$,

$p < 0.01$) was significantly negatively correlated with group performance. There was no significant correlation between collaborative performance and "marginal learner".

Table 7. The result of Spearman's correlation test

	Learner with Difficulties	Marginal Learner	Problem Solver	Coordinator	Leader
Grade	–.490**	–.294	.359*	.398*	.403*
Mean	1.13	1.84	1.26	.58	.29
SD	1.284	1.319	1.094	.620	.902

Notes: *p < .05, **p < .01.

To reveal the predictive ability of students with different roles, the study selected the "learner with difficulties", "problem solver", "coordinator" and "leader" as independent variables, which were significantly related to their performance, took performance as the dependent variable, and used multiple regression analysis to further analyze their predictive ability. The results are shown in Table 8.

Table 8. The results of multiple regression analysis

Independent Variables	β	t	Significance	R	R2	Adjusted R2	F	Significance
Learner with difficulties	–.152	–1.158	.257	.839[a]	.704	.659	15.487	.000[b]
Problem solver	.540	3.923	.001					
Coordinator	.454	3.989	.000					
Leader	.633	5.423	.000					

Notes: [a]Independent variables: Learner with difficulties, Problemsolver, Coordinationmanager, Leader [b]dependent variables: Grade.

Overall, $r2 = 0.704$, adjusted $r2 = 0.659$, $f = 15.487$, and $p < 0.05$, which indicates that there is a regression relationship between independent variables ("learner with difficulties", "problem solver", "coordinator" and "leader") and the dependent variable (collaborative performance). Independent variables can explain the variation of the dependent variable. From the significance of each independent variable to the dependent variable, "problem solver", "coordinator" and "leader" have significant influence on collaborative performance.

5 Discussion

The purpose of this study is to determine the emerging roles that group members play spontaneously without external intervention, and how the spontaneous roles of group members promote the success of the whole team, to better understand why groups with similar backgrounds produce significantly different quality performance results. This evidence is expected to promote a deep understanding of the nature of collaborative learning.

5.1 RQ1: Emerging role automatic detection

When no roles are assigned in advance, the contributions of individuals to the group and the interaction mode between individuals are generated in the process of interaction with other members of the group [9, 20]. In collaborative discussion, the task and the individual's understanding and inclination of the task will affect how students construct their collaboration, and form individual roles in collaborative learning activities and interaction with other team members [16].

For the first research question (RQ1), automatic identification of roles, we automatically recognize the discourse behavior of group members in collaborative discussion, and use the discourse behavior of the members as indicators to perform cluster analysis. The learners are clustered into five roles: "learner with difficulties", "marginal learner", "problem solver", "coordinator" and "leader".

From the results of cluster analysis, it can be seen that "leaders" perform best in almost all indicators. This kind of student can not only put forward new ideas and plans, but also actively express his or her own attitudes and opinions to other members, give reasons, organize and manage groups, and promote the learning atmosphere.

The students who take on the role of "problem solver" are more engaged in "statement" and "negotiation", and less engaged in "management" and "asking questions". This shows that this kind of student has his or her own views and opinions on the problems, expresses his or her own attitudes and reasons for the problems raised by others. These members contribute greatly to the problem solving of the group.

Students who play the role of "coordinator" are more involved in "management" and "negotiation". These students can manage and arrange team collaboration methods and time, and can propose feasible organizational plans, resource management and time management. Our research has also confirmed that these students are beneficial to good cooperation qualities. They can communicate with other members, understand the existing problems, and better plan and coordinate the collaborative process [24].

Students with the role of "marginal learner" have particularities in collaborative activities. Except for asking questions, this role is the worst among other indicators, indicating that these students hardly communicate with group members and rarely post. In this study, we found that these learners did not seem to actively interact with other members.

Students in the role of "learner with difficulties" are more engaged in "asking questions" and "sharing feelings", and less engaged in "statement" and "negotiation".

5.2 RQ2: Role composition of groups with different performances

For the second research question (RQ2), our research results show that the composition of participation roles played by group members is different between high-performance and low-performance groups, which is consistent with our expectations. The proportion of roles focused on problem-solving, especially those aiming at knowledge construction ("leader" and "problem solver"), is larger in the high-performing group than in the low-performing group. This result is in good agreement with the research results of Stempfle et al. on the effect of self-assigned roles in working groups [25]. In contrast, the role "learner with difficulties" is more common in low-performance groups than in high-performance groups.

However, in the research, we found that the number of “leader” is very small, and in some groups, there are even no “leaders” at all. In one group (Group 16), although there was a “leader”, the performance of the group was not satisfactory because there were three “learners with difficulties” in this group. This shows that high quality collaboration cannot rely on just a few excellent participants. Teachers should strive to ensure that every student participates in collaborative activities.

Overall, the quality of collaborative learning is clearly related to the composition of the participation roles of group members. This finding may have important implications for educators who use collaborative learning activities in teaching. Teachers can provide appropriate guidance and intervention according to the characteristics of each member’s speech behavior to promote group performance and results [24]. For example, teachers can provide guidance to students with learning difficulties, or encourage marginal learners to actively participate in discussions, answer questions or provide problem-solving suggestions, promote the transformation of their roles, and change the composition of group roles, to improve the quality of group collaboration.

5.3 RQ3: Impact of roles on group outcome

Regarding the third research question (RQ3), the impact of roles on group performance, our research results show that “problem solver”, “coordinator” and “leader” are significantly positively correlated with group collaboration performance, “learner with difficulties” is significantly negatively correlated with collaboration performance, and “marginal learner” is not correlated with performance (Table 5). Different roles have different contributions to problem solving and knowledge construction in collaborative discussion. On average, leaders have the greatest positive impact, which is consistent with the research of Zhang et al. and Hou [26, 27].

“Learners with difficulties” have the greatest negative impact on collaborative performance. Such students often ask questions in the discussion, and it is difficult for them to learn. At the same time, they rarely express their opinions on problem solving and management. When analyzing the discussion content of these learners, we found that there was no obvious deviation from the topic in their posts, indicating that the overall level of knowledge of such members was low. For these learners, teachers should pay more attention to their difficulties, and give more help and guidance to reduce their negative impacts on collaborative performance.

In contrast, “marginal learners” have the least influence on problem solving and knowledge construction. This kind of student does not actively participate in the discussion. However, by analyzing the content of their discussion, we found that some of their posts conveyed meaningful knowledge information and introduced new key terms, thereby promoting the final solution of the problem. Therefore, for these learners, teachers can encourage them to interact with other members, such as answering questions or providing suggestions, so as to promote their transition to the role of “problem solver”, thus positively affecting the performance of collaborative discussion.

6 Conclusion

This paper mainly studies how to automatically identify the spontaneous roles of group members in collaborative discussions, the differences in the role compositions of groups with different performances, and the impact of different roles on group outcomes. In terms of automatic role recognition, this paper constructs a text classification model to automatically identify the discourse intentions in collaborative discussions, and automatically recognize the roles of group members according to their discourse intentions, which overcomes the shortcomings of existing role recognition methods. Our research results provide preliminary evidence that under the computer-supported collaborative learning environment, through automatic analysis of the interactive content among participants, the emerging roles can be correctly identified. It not only provides a more comprehensive process-oriented participation role representation in theory, but also obtains the empirical verification in actual cases. In addition, the research uses a statistical analysis method to analyze the differences in role compositions of different collaborative performance groups, and finds that there are significant differences in the distributions of the "leader" role and the "learner with difficulties" role in high- and low-performance groups. Regarding the influence of roles on collaboration performance, our research results show that roles of the "leader", "problem solver" and "coordinator" have a positive influence on group performance, among which the role of "leader" has the greatest impact, while the "learners with difficulties" role has the greatest negative impact on group performance.

In short, this research provides comprehensive evidence that students play different roles spontaneously in collaborative learning and affect the effectiveness of collaborative learning, provides support and assistance for understanding the characteristics of collaborative learning and assists teachers in monitoring and improving the quality of collaborative learning. In this study, the number of courses involved in the experiment is small and the tasks are short, which limits the universality of the research results. In future research, we will increase the number of courses and the difficulty of tasks to gain further insights. In future research, we also plan to explore the evolution of the role model of the same group of students in different tasks, and analyze the influence of students' individual knowledge level, preference for cooperative behavior and emotional factors on their roles.

7 Acknowledgment

This work is partially supported by the 14th Five Year Plan project for Educational Science in Henan Province (No. 2021YB0124) and Henan Science and Technology project (No. 232102321062). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

8 References

[1] Zheng, Y. F., Xu, C., Su, Y. (2018). Measuring and visualizing group knowledge elaboration in online collaborative discussions. Educational Technology & Society, 21: 91–103.

[2] Galikyan, I., Admiraal, W. (2019). Students' engagement in asynchronous online discussion: The relationship between cognitive presence, learner prominence, and academic performance. The Internet and Higher Education, 43: 100692.1–100692.9. https://doi.org/10.1016/j.iheduc.2019.100692

[3] Volet, S., Vauras, M., Salo, A. E., Khosa, D. (2017). Individual contributions in student-led collaborative learning: Insights from two analytical approaches to explain the quality of group outcome. Learning & Individual Differences, 53: 79–92. https://doi.org/10.1016/j.lindif.2016.11.006

[4] Nia M. M. Dowell, Lin, Y.W., Andrew Godfrey, Christopher Brooks. (2020). Exploring the relationship between emergent socio-cognitive roles, collaborative problem-solving skills, and outcomes: A group communication analysis. Journal of Learning Analytics, 7: 38–57. https://doi.org/10.18608/jla.2020.71.4

[5] Ouyang, F., Chang, Y. H. (2019). The relationships between social participatory roles and cognitive engagement levels in online discussions. British Journal of Educational Technology, 50: 1396–1414. https://doi.org/10.1111/bjet.12647

[6] Strijbos, J. W., Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. Computers in Human Behavior, 26: 491–494. https://doi.org/10.1016/j.chb.2009.08.006

[7] Capuano, N., Mangione, G. R., Mazzoni, E., Miranda, S., Orciuoli, F. (2014). Wiring role taking in collaborative learning environments. SNA and Semantic Web can improve CSCL script? International Journal of Emerging Technologies in Learning (iJET), 9(7): 30–38. https://doi.org/10.3991/ijet.v9i7.3719

[8] Dowell, N. M., Nixon, T., Graesser, A. C. (2018).Group communication analysis: A computational linguistics approach for detecting socio-cognitive roles in multi-party interactions. Behavior Research Methods, 51: 1007–1041. https://doi.org/10.3758/s13428-018-1102-z

[9] De Laat, M., Lally, V. (2004). It's not so easy: Researching the complexity of emergent participant roles and awareness in asynchronous networked learning discussions. Journal of Computer Assisted Learning, 20: 165–171. https://doi.org/10.1111/j.1365-2729.2004.00085.x

[10] Mudrack, P. E., Farrell, G. M. (1995). An examination of functional role behaviour and its consequences for individuals in group settings. Small Group Research, 26: 542–571. https://doi.org/10.1177/1046496495264005

[11] Brush, T. A. (1998). Embedding cooperative learning into the design of integrated learning systems: Rationale and guidelines. Educational Technology Research & Development, 46: 5–18. https://doi.org/10.1007/BF02299758

[12] Tawafak, R. M., Alyoussef, I. Y. . ., Al-Rahmi, W. M. (2023). Essential factors to improve student performance using an e-learning model: Review study. International Journal of Interactive Mobile Technologies (iJIM), 17(03): 160–176. https://doi.org/10.3991/ijim.v17i03.35727

[13] Shestak, V., Gura, A., Borisova, U., Kozlovskaya, S. (2021). The role of social networks in the organization of the educational process and learning. International Journal of Interactive Mobile Technologies (iJIM), 15(11): 96–112. https://doi.org/10.3991/ijim.v15i11.21545

[14] Stewart, G. L., Fulmer, I. S., Barrick, M. R. (2005). An exploration of member roles as a multilevel linking mechanism for individual traits and team outcomes. Personnel Psychology, 58: 343–365. https://doi.org/10.1111/j.1744-6570.2005.00480.x

[15] Strijbos, J. W., Laat, M. (2010). Developing the role concept for computer-supported collaborative learning: An explorative synthesis. Computers in Human Behaviour, 26: 495–505. https://doi.org/10.1016/j.chb.2009.08.014
[16] Fischer, F., Kollar, I., Mandl, H., Haake, J. M. (2007). Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives. Springer: New York, USA. https://doi.org/10.1007/978-0-387-36949-5
[17] Dragan, G., Srećko, J., Eagan, B. R., Williamson, S. D. (2018). SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. Computers in Human Behavior, 92: 562–577. https://doi.org/10.1016/j.chb.2018.07.003
[18] Marcos-García, J. A., Martínez-Mones, A., Dimitriadis, Y., Anguita-Martínez, R.(2007). A role-based approach for the support of collaborative learning activities. e-Service Journal. Special Issue: Collaboration and Interaction in Varied Environments, 6(1): 89–112. https://doi.org/10.2979/esj.2007.6.1.40
[19] Marcos-García, J.-A., Martínez-Monés, A., Dimitriadis, Y. (2015). DESPRO: A method based on roles to provide collaboration analysis support adapted to the participants in CSCL situations. Computers & Education, 82: 335–353. https://doi.org/10.1016/j.compedu.2014.10.027
[20] De Laat, M., Lally, V., Lipponen, L., et al. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for social network analysis. Computer Supported Learning, 2: 87–103. https://doi.org/10.1007/s11412-007-9006-4
[21] Cesareni, D., Cacciamani, S., Fujita, N. (2015). Role taking and knowledge building in a blended university course. International Journal of Computer-Supported Collaborative Learning, 11: 9–39. https://doi.org/10.1007/s11412-015-9224-0
[22] Hou, H. T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. Computers in Human Behavior, 48: 424–435. https://doi.org/10.1016/j.chb.2015.02.010
[23] Lehmann-Willenbrock, N., Beck, S. J., Kauffeld, S. (2016). Emergent team roles in organizational meetings: Identifying communication patterns via cluster analysis. Communication Studies, 67: 37–57. https://doi.org/10.1080/10510974.2015.1074087
[24] Xu, C., Zheng, Y. F., Hu, H. N. (2016). Exploring individual contributions during online collaborative discussions. Journal of Computers in Education, 3: 395–411. https://doi.org/10.1007/s40692-016-0071-4
[25] Stempfle, J., Hubner, O., Badke-Schaub, P. (2001). A functional theory of task role distribution in work groups. Group Processes & Intergroup Relations, 4(2): 138–159. https://doi.org/10.1177/1368430201004002005
[26] Zhang, S., Liu, Q., Chen, W., Wang, Q., Huang, Z. (2017). Interactive networks and social knowledge construction behavioural patterns in primary school teachers' online collaborative learning activities. Computers & Education, 104: 1–17. https://doi.org/10.1016/j.compedu.2016.10.011
[27] Hou, H. (2015). What makes an online community of practice work? A situated study of Chinese student teachers' perceptions of online professional learning. Teaching and Teacher Education, 46: 6–16. https://doi.org/10.1016/j.tate.2014.10.005

9 Authors

Huichao Mi is a lecturer at College of Computer and Information Engineering, Henan University of Economics and Law, China. She received her master's degree from Harbin Engineering University in 2005. Her research focuses on education data mining and analysis of online learning processes.

Qiaorong Zhang is a professor at College of Computer and Information Engineering, Henan University of Economics and Law, China. Her research interests include learning analysis technology, education data mining, and AI in education.

Yafeng Zheng is a professor at Institute of Advanced Studies in Humanities and Social Sciences, Beijing Normal University at Zhuhai. Her expertise is learning analysis, data mining and big data visualization in education.

Article submitted 2023-02-09. Resubmitted 2023-03-27. Final acceptance 2023-03-27. Final version published as submitted by the authors.