Investigating Conversational Agent Action in Legal Case Retrieval

Bulou Liu[1*], Yiran Hu[2*], Yueyue Wu[1], Yiqun Liu[1⋆], Fan Zhang[3], Chenliang Li[4], Min Zhang[1], Shaoping Ma[1], and Weixing Shen[2]

[1] Department of Computer Science and Technology, Institute for Internet Judiciary, Tsinghua University. Quan Cheng Laboratory.
[2] School of Law, Tsinghua University.
[3] School of Information Management, Wuhan University.
[4] School of Cyber Science and Engineering, Wuhan University.
yiqunliu@tsinghua.edu.cn

Abstract. Legal case retrieval is a specialized IR task aiming to retrieve supporting cases given a query case. Existing work has shown that the conversational search paradigm can improve users' search experience in legal case retrieval with humans as intermediary agents. To move further towards a practical system, it is essential to decide what action a computer agent should take in conversational legal case retrieval. Existing works try to finish this task through Transformer-based models based on semantic information in open-domain scenarios. However, these methods ignore search behavioral information, which is one of the most important signals for understanding the information-seeking process and improving legal case retrieval systems. Therefore, we investigate the conversational agent action in legal case retrieval from the behavioral perspective. Specifically, we conducted a lab-based user study to collect user and agent search behavior while using agent-mediated conversational legal case retrieval systems. Based on the collected data, we analyze the relationship between historical search interaction behaviors and current agent actions in conversational legal case retrieval. We find that, with the increase of agent-user interaction behavioral indicators, agents are increasingly inclined to return results rather than clarify users' intent, but the probability of collecting candidates does not change significantly. With the increase of the interactions between the agent and the system, agents are more inclined to collect candidates than clarify users' intent and are more inclined to return results than collect candidates. We also show that the agent action prediction performance can be improved with both semantic and behavioral features. We believe that this work can contribute to a better understanding of agent action and useful guidance for developing practical systems for conversational legal case retrieval.

Keywords: Conversational search, agent action, legal case retrieval

[*] Equal contributions from both authors.
[⋆] Corresponding author.

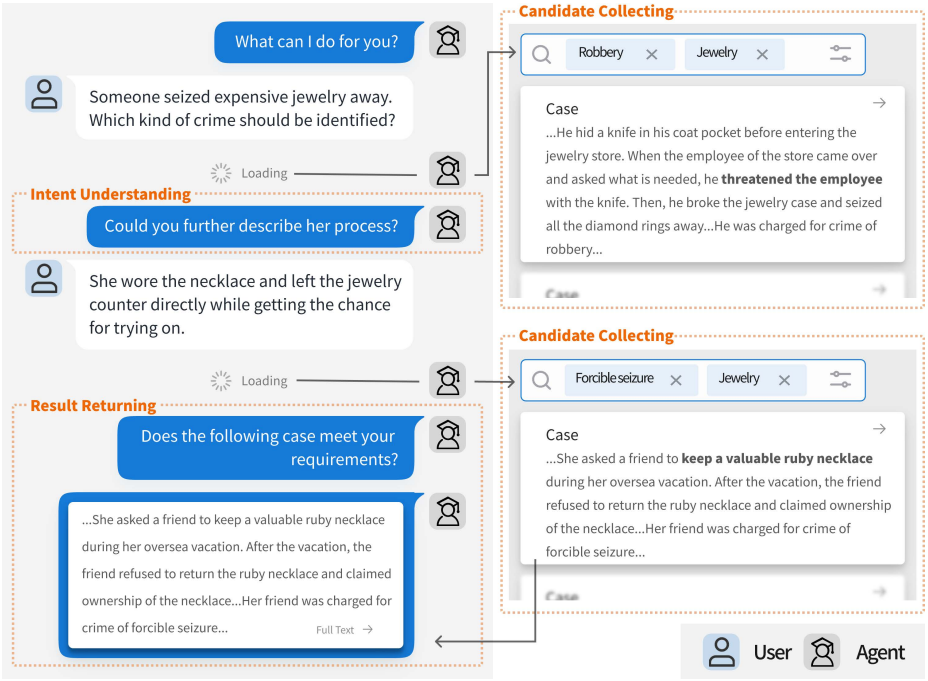

Fig. 1. An example of the three kinds of agent actions in conversational legal case retrieval.

1 Introduction

In recent years, legal case retrieval has attracted much attention in the IR research community. It aims to retrieve supporting cases for a given query case and contributes to better legal systems. Existing works show that an automatic system not only performs the legal case retrieval tasks with higher performance than lawyers, but also performs more efficiently [15]. Under traditional legal case retrieval systems, users need to issue queries to express their complex information needs [6, 14], which requires sufficient domain knowledge [19, 12, 23]. Liu et al. [8, 9] show that conversational search paradigm, where human experts play the role of intermediary conversational agents, can be adopted to improve users' legal case retrieval experience in terms of query formulation, result examination, and users' satisfaction and search success.

The conversational agent action prediction task aims to decide what action the agents will take based on the context of the conversation and helps provide useful and meaningful responses in conversational search systems [2]. As shown in Figure 1, there are three kinds of conversational agent actions in legal case retrieval:

– **Intent Understanding (IU):** The agents ask clarifying questions to understand users' search intent better.

- **Candidate Collecting (CC):** The agents submit queries to a legal case ad-hoc search system and collect candidate cases.
- **Result Returning (RR):** The agents select relevant cases from candidates and return them to users as results.

It is essential to understand human conversational agent action and decide what action to take automatically before we move further towards a practical conversational search system (that is, to use an automated agent instead of a human expert) for legal case retrieval. Existing works try to solve the conversational agent action prediction problem through Transformer-based models [10, 5] which exploit semantic information in open-domain scenarios [22]. However, behavioral information, which is one of the most important signals for understanding the information-seeking process [3] and providing implicit feedback for legal case retrieval system [19], has not been incorporated into conversational agent action prediction in legal case retrieval.

This paper investigates the conversational agent action in legal case retrieval from the behavioral perspective. Different from traditional search systems, conversational search systems contain two kinds of behavioral information (i.e., user and agent behaviors). Specifically, we analyze the relationship between the historical interaction behaviors and the current agent actions in legal case retrieval from two aspects: agent-user interactions and agent-system interactions. Furthermore, we try to utilize behavioral features to predict which action the agent would like to take. Our research questions are as follows:

- **RQ1:** What is the relationship between the historical behaviors and the current agent actions in legal case retrieval?
- **RQ2:** Can we improve the conversational agent action prediction performance with behavioral features involved in legal case retrieval?

To shed light on these research questions, we conducted a lab-based user study with 106 tasks to collect user and agent search behavior using agent-mediated conversational legal case retrieval systems. It's worth noting that no available conversational legal case retrieval system exists currently. Therefore, we recruit legal experts as intermediary agents to complete the procedure in a wizard-of-oz fashion. To answer RQ1, we compare the differences in user and agent historical behaviors w.r.t. different conversational agents' actions. Furthermore, we define the conversational agent action prediction as a classification task and demonstrate the effectiveness of features extracted from the user and agent behaviors for RQ2.

2 Related work

Legal case retrieval is a specialized IR task that differs from general web search in various aspects, such as the needs for data authority [1] and the definition of relevance [6, 7, 18, 11]. Behavior information plays an important role in legal case retrieval. [19] investigated user behavior in legal case retrieval. They found

that users of legal case retrieval devote more search effort and appear to be more patient and cautious. They further applied the behavioral features to relevance prediction. [8] have shown that conversational search paradigm can be adopted to improve users' legal case retrieval experience in terms of query formulation, result examination, and users' satisfaction and search success. They revealed that it is necessary to develop conversational legal case retrieval systems.

And agent action prediction is important for develop practical conversational systems. [16] proposed a hierachical deep reinforcement learning approach to learning the dialogue policy at different temporal scales. [21] presented an agent that efficiently learns dialogue policy from demonstrations through policy shaping and reward shaping. [22] proposed a Transformer-based model to predict agent action in conversational search systems in open-domain scenarios.

Compared to these studies, our work focuses on the behavioral perspective of the conversational agent action in legal case retrieval.

3 User Study

To investigate the relationship between the historical behaviors and the current agent actions in conversational legal case retrieval, we conducted a lab-based user study with 106 tasks. We show the details in this section.

3.1 Conversational Legal Case Retrieval

It's worth noting that no available conversational legal case retrieval system exists currently. We add an intermediary agent to complete the procedure in a wizard-of-oz fashion. The agent needs to understand users' intents via conversations, construct queries and pick cases from SERPs for the user. Specifically, the procedure contains the following steps:

1. The user submits a legal issue question to the agent in natrual language.
2. The agent asks clarifying questions until the background information of the search issue is sufficient.
3. The agent submits queries to the legal case retrieval system. She then selects cases from the SERPs and responds to the user with the selected ones.

In particular, the conversational legal case retrieval procedure contains rich logs of behavioral information. On the one hand, it contains agents' interactions with users, such as search questions, clarifying questions, and the cases returned by the agent. On the other hand, it includes agents' interactions with the system, such as queries and clicks. Therefore, we extracted ten behavioral features from two aspects: agent-user interactions and agent-system interactions, which are shown in Table 1. In detail, as for agent-user interaction behaviors, we focus on conversational input behaviors and agent answering behaviors. As for agent-system interactions, we concentrate on query formulation and the search engine result page (SERP) examination behaviors, especially the examination behaviors in the SERPs from the last query. Note that there were no users' interactions

with the system because the user study dataset was collected in a wizard-of-oz approach. And we just kept behavioral information before the current action for analysis, i.e., the same setting for the action prediction task.

Table 1. The list of 10 behavioral features extracted from the agent-user interactions and agent-system interactions.

Group	Behavioral Features
Agent-User	Number of utterances/words in conversations
	Number of returned results/returned cases
Agent-System	Number of queries/query words
	Number of clicks in all queries/in last query
	Avg./Max. click rank in last query

3.2 Tasks and Participants

We collected 106 search tasks from legal practitioners' real information need via online forums and social networks, covering 3 legal domains: 34 civil tasks (involving 10 topics, such as "Inheritance", "Personality rights", "Contracts" and "Marriage"), 35 criminal tasks (involving 7 topics, such as "Robbery", "Fraud", "Bribery", "Forcible Rape" and "Traffic accident") and 37 commercial tasks (involving 9 topics, such as "Company", "Expertise Bankruptcy" and "Insurance"). Compared with existing user studies for legal case retrieval or conversational search [19, 20], we believe that the number of tasks is enough for a between-subjects analysis. Each task contained a query case description and a legal issue. Users were expected to retrieve legal cases which may help to answer the issue question.

There were two kinds of participants: users and agents. As for users, we recruited 30 participants (12 males and 18 females) via online forums and social networks. They were all native Chinese speakers and college law students. All users had no previous experience with conversational search systems. No users conducted two tasks in the same topic, which also can avoid the task learning effects on the results. Note that the tasks have negligible or no learning effects on each other even in the same domain if they are not in the same topics.

We recruited 15 graduate students from law school (5 for civil law, 5 for criminal law and 5 for commercial law) to be agents. They were all native Chinese speakers and qualified in legal practice[5]. To ensure an adequate level of domain expertise, they only participated in the task related to their research fields. In addition, they all achieved a score of 95 or more in the courses corresponding to their experimental topics. This can reduce the effect of individual variability. And they were trained with 5 auxiliary search tasks beforehand to familiarize with the query construction skills in the legal case retrieval system, guaranteeing an adequate level of search expertise.

[5] They had passed the "National Uniform Legal Profession Qualification Examination"

Table 2. Statistics of agent actions in the user study dataset.

#Tasks	#Intent Understanding	#Candidate Collecting	#Result Returning
106	437	385	208

As for the legal case retrieval system, we choose a leading commercial legal search engine[6] in China. Users and agents had a conversation (just in text form) via Zoom[7].

3.3 Procedure

Before the experiments, we firstly requested each participant to complete a warm-up search task. We then introduce the details of the procedure as follows:

Query Case and Issue Reading. In the first step, the user read the query case description and the legal issue carefully. She could refer to the query case at any time during the session, so she did not need to memorize the case description at this step.

Pre-task Questionnaire. Next, the user was asked to finish a pre-search questionnaire, including: domain knowledge level, task difficulty level, and prior interest level of the task with a 5-point Likert scale (1: not at all, 2: slightly, 3: somewhat, 4: moderately, 5: very).

Task Completion. After that, the user started performing searches with the agent. At this step, we collected the agent's interactions with the system, including queries, clicks, etc. Moreover, we recorded the conversation contents, including users' legal questions, agents' clarifying questions, the cases returned by the agent.

Post-task Questionnaire. After examining the supporting cases returned by the agent, the user was required to complete a post-task questionnaire. At this step, we collected explicit feedback signals with respect to the search experience, including five-grade workload and satisfaction.

Result Assessment. After completing the post-task questionnaire, the user was further asked to annotate the cases that agents clicked in the SERPs. That is, a relevance score is annotated to each case (1: irrelevant, 2: relevant). As for the cases that weren't clicked, we simply regarded them as irrelevant.

To drive the conversational legal case retrieval process, the intermediary agents can take three kinds of actions (i.e., Intent Understanding, Candidate Collecting, and Result Returning). Through these actions, the agents understand user intent by clarifying questions, collect candidate cases from the traditional legal search system by submitting queries, and return relevant cases from candidates to users, respectively. Table 2 shows distribution of each agent action in the user study dataset.

[6] https://ydzk.chineselaw.com/case
[7] https://zoom.us/

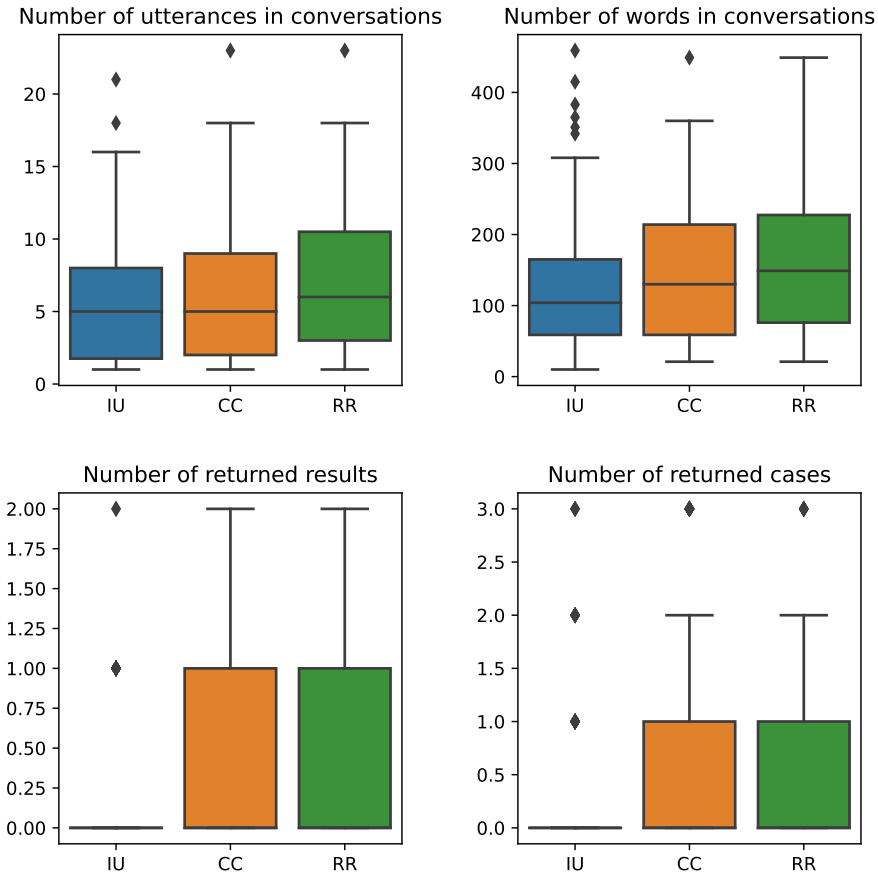

Fig. 2. Comparison of historical agent-user interaction behavioral measures given different current agent actions.

4 Results

4.1 Analysis on Conversational Agent Action

To address **RQ1**, we report the relationship between the historical interaction features and the current agent action using box plots. Specifically, we compare the differences in users' and agents' historical behaviors given different conversational agents' actions from two aspects: agent-user interactions and agent-system interactions. We also perform a series of one-way ANOVA tests [4] and pairwise t-tests [17] to determine the significance.

Comparison of agent-user interactions. Firstly, we compare historical agent-user interaction behaviors w.r.t. different agent actions. Here, we focus on conver-

sational input behaviors and agent answering behaviors. The results of ANOVA tests (ANOVA-p) are reported in Figure 2. We can make the following observations.

From the first line in Figure 2, we can observe that the conversational input behavioral indicators (i.e., the number of utterances and words) show significant differences between the three agent actions (ANOVA-$p < 0.05$). Moreover, we find that the number of utterances and words under "Result Returning" action is significantly more than that under "Intent Understanding" action ($p < 0.05$). However, there are no significant differences according to pairwise t-tests in conversational input behaviors between the "Candidate Collecting" action and the other actions. This illustrates that the agent tends to adopt the "Intent Understanding" action when the conversation length is short and tends to adopt the "Result Returning" action when the conversation contains sufficient information. Furthermore, the agents may take the "Candidate Collecting" action regardless of the length of the conversation.

We further investigate the agent answering behaviors (i.e., the number of returned results and returned cases), and the results of ANOVA tests are shown in the second line in Figure 2. There are also significant differences in these two indicators before agents adopt the three actions (ANOVA-$p < 0.01$). Furthermore, we find that the number of returned results and returned cases before agents take the "Intent Understanding" action are significantly less than that before agents take the other two actions ($p < 0.01$). And these two indicators do not show significant differences before the "Candidate Collecting" action and the "Result Returning" action. These indicate that as the agent answering behavioral indicators increases, the agent will decrease the probability of taking the "Intent Understanding" action and prefer to take the other two actions.

Comparison of agent-system interactions. Then we compare historical agent-system interaction behaviors w.r.t. different agent actions. Specifically, we concentrate on query formulation and SERP examination behaviors. The results of ANOVA tests (ANOVA-p) are reported in Figure 3. We can make the following observations.

Overall, we find that the historical agent-system interaction behavioral indicators before the three conversational agent actions follow the following relative order: "Intent Understanding" < "Candidate Collecting" < "Result Returning" (shown in Figure 3). We can observe that agents submitted fewer queries and query words before "Intent Understanding" actions than those before another action ($p < 0.001$). And the number of queries and query words before taking "Candidate Collecting" actions are less than those before taking "Result Returning" actions ($p < 0.001$). The above phenomenons also exist for the number of clicks, especially in the last query. This suggests that with the increase of the interactions between the agent and the system, agents are more inclined to collect candidates than clarify users' intent and are more inclined to return results than collect candidates.

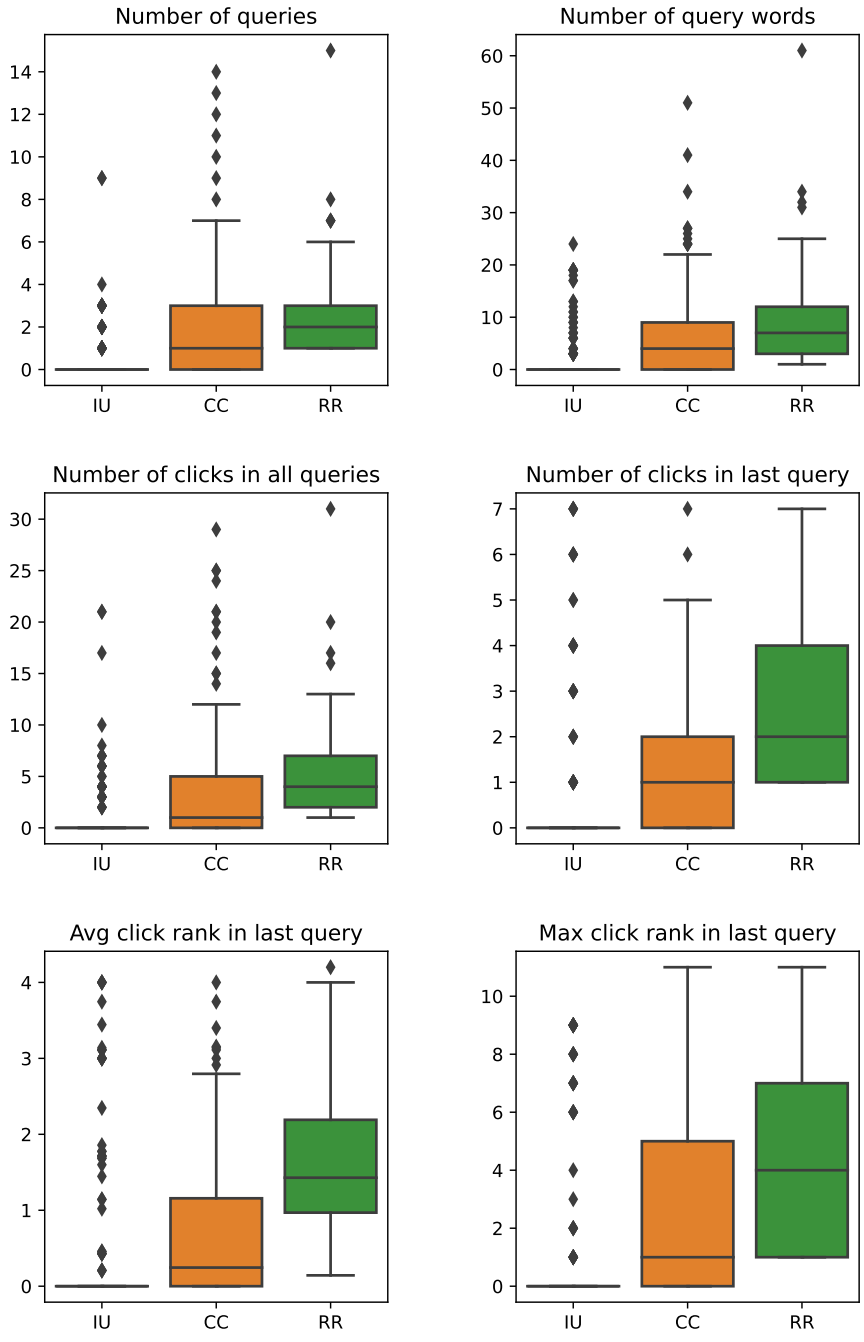

Fig. 3. Comparison of historical agent-system interaction behavioral measures given different current agent actions.

Table 3. Performance comparison of conversational agent action prediction task. AU and AS denotes that the method incorporates agent-user and agent-system behavioral features, respectively. Best results are in boldface. † indicates that the difference to Random is statistically significant at 0.05 level from the student t-test.

Method	Overall	IU vs. CC	IU vs. RR	CC vs. RR
Random	0.3607	0.5259	0.5158	0.5112
AU	0.4018	0.5625	$0.8087^{\dagger}$	0.6182
AS	$0.5142^{\dagger}$	$0.7775^{\dagger}$	$0.8171^{\dagger}$	$0.6971^{\dagger}$
AU+AS	$\mathbf{0.5265}^{\dagger}$	$\mathbf{0.8058}^{\dagger}$	$\mathbf{0.8260}^{\dagger}$	$\mathbf{0.7043}^{\dagger}$

Furthermore, we investigate two indicators related to the examination behaviors in the last SERP: the average/maximum click rank in the last query. The results of ANOVA tests are shown in the last line in Figure 3. Similarly, these two indicators show significant differences between the three actions (ANOVA-$p < 0.001$). And we can observe that the click rank is larger before the "Result Returning" action than that before the other two actions significantly ($p < 0.001$). It shows that the agents are more inclined to return results when they examine and click on the results with a larger rank. Because they are more convinced that they have understood user intent and submitted accurate queries.

Summary. To answer **RQ1**, our findings are as follows: 1) With the increase of agent-user interaction behavioral indicators, agents are increasingly inclined to return results rather than clarify users' intent, but the probability of collecting candidates does not change significantly; 2) With the increase of the interactions between the agent and the system, agents are more inclined to collect candidates than clarify users' intent and are more inclined to return results than collect candidates.

4.2 Conversational Agent Action Prediction

To address **RQ2**, we try to improve the conversational agent action prediction with behavioral features. Two groups of features are adopted in the experiments: agent-user behaviors and agent-system behaviors (ref. Table 1). We define the prediction task as a multi-class classification task and use Macro-F1 for evaluation. Furthermore, we further analyze the effect of these features through three binary classification tasks: IU vs. CC, IU vs. RR and CC vs. RR. We use the F1-score to evaluate the three classification tasks.

We first investigate the prediction performance without semantic features. Random is the baseline which decides actions randomly. As this task can be treated as a multi-class classification problem, we apply a gradient boosting classifier [13] and perform 5-fold cross-validation. The results are shown in Table 3. We can observe that using both groups of features achieves the best performance on all classification tasks and using agent-system interaction behavioral features also outperforms Random significantly. These suggest that both groups

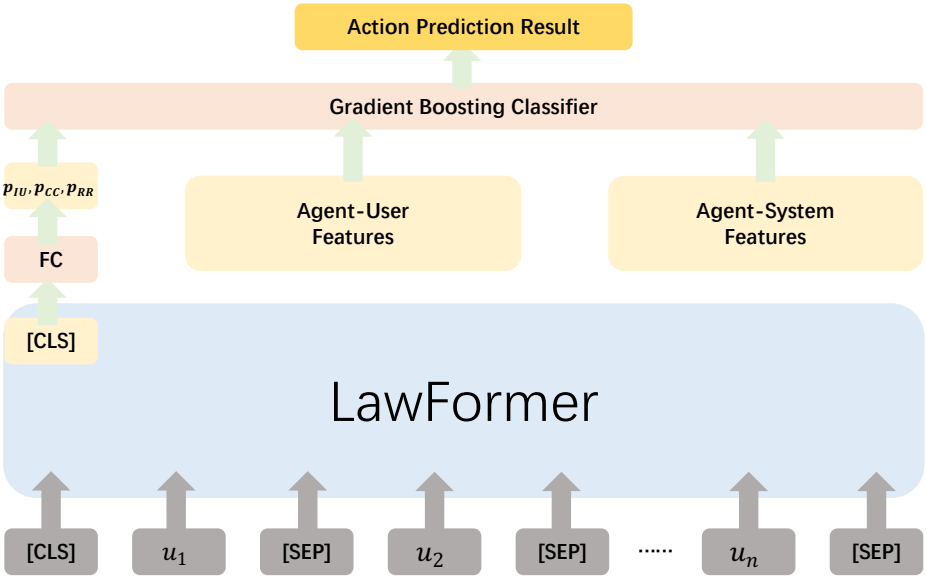

Fig. 4. Combining semantic and behavioral features for the conversational agent action prediction. $u_1, u_2, ..., u_n$ denote all the utterances in the conversation. *FC* denotes the full connected layers. p_{IU}, p_{CC}, p_{RR} denote the probabilities of taking the three actions predicted by the LawFormer baseline.

of features are useful for the conversational agent action prediction in legal case retrieval. And agent-user behaviors only significantly improve the performance of IU vs. RR task, suggesting that they do not provide much information to distinguish whether to take the Candidate Collecting action.

Existing works try to solve the conversational agent action prediction problem through Transformer-based models based on semantic information. To further investigate the prediction performance with semantic features, we concatenated all the utterances in the conversation together and used LawFormer [24] as the encoder. Here LawFormer is a Longformer-based pre-trained language model for Chinese legal long documents understanding. Then we fed the [CLS] embedding into full connected layers and fine-tuned LawFormer for each tasks as the baseline. The model is optimized by the cross-entropy loss. We performed 5-fold cross-validation and the results are shown in Table 4. We can find LawFormer outperforms all the methods without involving semantic features (shown in Table 3). It illustrates that semantic information is also very useful for the conversational agent action prediction in legal case retrieval.

Then we regarded the probability distribution (i.e., the probabilities of taking the three actions, 3-dimensional in total) of the full connected layers' output as semantic features. We combine the semantic features with behavioral features together and also apply a gradient boosting classifier to obtain the final agent action prediction results. Note that we just utilize the LawFormer fine-

Table 4. Performance comparison of conversational agent action prediction task with using semantic features. AU and AS denotes that the method incorporates agent-user and agent-system behavioral features, respectively. Best results are in boldface. † indicates that the difference to LawFormer is statistically significant at 0.05 level from the student t-test.

Method	Overall	IU vs. CC	IU vs. RR	CC vs. RR
LawFormer	0.5425	0.8232	0.8739	0.7335
LawFormer+AU	0.5675	0.8318	0.8844	0.7398
LawFormer+AS	0.5828	0.8428	0.8462	0.7564
LawFormer+AU+AS	**0.6177**†	**0.8669**†	**0.8870**	**0.8015**†

tuned in the baseline and do not take further fine-tuning strategies. And the model framework is shown in Figure 4. As for the three binary classification tasks: IU vs. CC, IU vs. RR and CC vs. RR, the semantic features change from three-dimensional to two-dimensional as the output of the fully connected layer changes. Other experimental settings remain the same and the results are shown in Table 4. AU and AS denotes that the method incorporates agent-user and agent-system behavioral features, respectively. We find that combining behavioral features with semantic features achieves significantly better classification performance than LawFormer, especially in the IU vs. CC task and the CC vs. RR task. This illustrates that historical behaviors are useful supplementary information for semantic features to distinguish whether to take the Candidate Collecting action.

Concerning **RQ2**, we find that behavioral features can improve the conversational agent actions prediction performance in legal case retrieval whether semantic features are involved or not.

5 Conclusion

In this paper, we investigate three kinds of conversational agent actions (i.e., Intent Understanding, Candidate Collecting, and Result Returning) in legal case retrieval from a behavioral perspective. We find that with the increase of agent-user interaction behavioral indicators, agents are increasingly inclined to return results rather than clarify users' intent, but the probability of collecting candidates does not change significantly. Moreover, with the increase of the interactions between the agent and the system, agents are more inclined to collect candidates than clarify users' intent and are more inclined to return results than collect candidates. We further show that the agent action prediction performance can be improved with both semantic and behavioral features in legal case retrieval. We believe that this work can contribute to a better understanding of agent action and useful guidance for developing practical systems for conversational legal case retrieval.

As for future work, we firstly plan to utilize more sophisticated algorithms (e.g., reinforcement learning) to incorporate behavioral information into the legal

conversational agent action prediction task more effectively. Secondly, we prepare to take more fine-grained behavioral information (e.g., mouse movements, hovers and so on) into consideration. At last, we also try to improve retrieval performance through more accurate action prediction in conversational legal case retrieval.

Acknowledgement

This work is supported by the Natural Science Foundation of China (Grant No. 61732008, 62002194) and Tsinghua University Guoqiang Research Institute.

References

1. Arewa, O.B.: Open access in a closed universe: Lexis, westlaw, law schools, and the legal information market. Lewis & Clark L. Rev. **10**, 797 (2006)
2. Azzopardi, L., Dubiel, M., Halvey, M., Dalton, J.: Conceptualizing agent-human interactions during the conversational search process. In: The second international workshop on conversational approaches to information retrieval (2018)
3. Buscher, G., Dengel, A., Biedert, R., Elst, L.V.: Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. ACM Transactions on Interactive Intelligent Systems (TiiS) **1**(2), 1–30 (2012)
4. Cuevas, A., Febrero, M., Fraiman, R.: An anova test for functional data. Computational statistics & data analysis **47**(1), 111–122 (2004)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
6. Ferrer, A.S., Hernández, C.F., Boulat, P.: Legal search: foundations, evolution and next challenges. the wolters kluwer experience. Revista Democracia Digital e Governo Eletrônico **1**(10), 120–132 (2014)
7. Fleiss, J.L.: Measuring nominal scale agreement among many raters. Psychological bulletin **76**(5), 378 (1971)
8. Liu, B., Wu, Y., Liu, Y., Zhang, F., Shao, Y., Li, C., Zhang, M., Ma, S.: Conversational vs traditional: Comparing search behavior and outcome in legal case retrieval. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1622–1626 (2021)
9. Liu, B., Wu, Y., Zhang, F., Liu, Y., Wang, Z., Li, C., Zhang, M., Ma, S.: Query generation and buffer mechanism: Towards a better conversational agent for legal case retrieval. Information Processing & Management **59**(5), 103051 (2022)
10. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
11. Ma, Y., Shao, Y., Liu, B., Liu, Y., Zhang, M., Ma, S.: Retrieving legal cases from a large-scale candidate corpus. Proceedings of the Eighth International Competition on Legal Information Extraction/Entailment, COLIEE2021 (2021)
12. Mao, J., Liu, Y., Kando, N., Zhang, M., Ma, S.: How does domain expertise affect users' search interaction and outcome in exploratory search? ACM Transactions on Information Systems (TOIS) **36**(4), 1–30 (2018)

13. Mason, L., Baxter, J., Bartlett, P., Frean, M.: Boosting algorithms as gradient descent. Advances in neural information processing systems **12** (1999)
14. McGinnis, J.O., Pearce, R.G.: The great disruption: How machine intelligence will transform the role of lawyers in the delivery of legal services. Actual Probs. Econ. & L. p. 1230 (2019)
15. McGinnis, J.O., Wasick, S.: Law's algorithm. Fla. L. Rev. **66**, 991 (2014)
16. Peng, B., Li, X., Li, L., Gao, J., Celikyilmaz, A., Lee, S., Wong, K.F.: Composite task-completion dialogue policy learning via hierarchical deep reinforcement learning. arXiv preprint arXiv:1704.03084 (2017)
17. Semenick, D.: Tests and measurements: The t-test. Strength & Conditioning Journal **12**(1), 36–37 (1990)
18. Shao, Y., Liu, B., Mao, J., Liu, Y., Zhang, M., Ma, S.: Thuir@ coliee-2020: Leveraging semantic understanding and exact matching for legal case retrieval and entailment. arXiv preprint arXiv:2012.13102 (2020)
19. Shao, Y., Wu, Y., Liu, Y., Mao, J., Zhang, M., Ma, S.: Investigating user behavior in legal case retrieval. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 962–972 (2021)
20. Vtyurina, A., Savenkov, D., Agichtein, E., Clarke, C.L.: Exploring conversational search with humans, assistants, and wizards. In: Proceedings of the 2017 chi conference extended abstracts on human factors in computing systems. pp. 2187–2193 (2017)
21. Wang, H., Peng, B., Wong, K.F.: Learning efficient dialogue policy from demonstrations through shaping. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 6355–6365 (2020)
22. Wang, Z., Ai, Q.: Controlling the risk of conversational search via reinforcement learning. In: Proceedings of the Web Conference 2021. pp. 1968–1977 (2021)
23. White, R.W., Dumais, S.T., Teevan, J.: Characterizing the influence of domain expertise on web search behavior. In: Proceedings of the second ACM international conference on web search and data mining. pp. 132–141 (2009)
24. Xiao, C., Hu, X., Liu, Z., Tu, C., Sun, M.: Lawformer: A pre-trained language model for chinese legal long documents. AI Open **2**, 79–84 (2021)