Deep Reinforcement Learning-Based Automated Negotiation Model for Shared Decision-Making

Sijie Wen, Zhongqi Cai, Xin Chen, Rende Hong, Kaibiao Lin, Ping Lu

***Abstract*—Shared Decision-Making (SDM) is a crucial approach in clinical practice that aims to integrate patient preferences with medical expertise. However, existing SDM approaches often rely on artificial processes and fixed strategies, which limit their ability to adapt to nuanced preferences and diverse concession patterns in complex medical decision scenarios. To address these limitations, we propose the Actor-Critic (AC) Network-Based Agent Negotiation Model for SDM (ACANM-SDM) by employing agent-based technology and adaptive negotiation strategies to automate the SDM process. The ACANM-SDM consists of three key components: (1) a preference representation module using trapezoidal fuzzy affiliation functions, (2) a pre-training data generation mechanism based on issue weights, and (3) dual AC networks for training adaptive offer and acceptance strategies. Specifically, the preference module captures doctor-patient preferences, which serve as the pre-training data inputs to the AC networks. These networks then generate and evaluate negotiation strategies, adapting to individual concession patterns. Finally, the generated strategy is applied to the doctor-patient agent and the alternating bidding protocol is used to build the SDM automatic negotiation model. Experimental results on real-world clinical datasets demonstrate that ACANM-SDM outperforms four state-of-the-art methods in shared decision-making tasks. Notably, ACANM-SDM achieves a 26% improvement in social welfare utility compared to the best-performing baseline. Furthermore, ACANM-SDM enhances efficiency by reducing the average number of negotiation rounds by 50%.**

***Index Terms*—Shared Decision-Making, Deep Reinforcement Learning, Actor-Critic Network, Automated Negotiation, Multi-Agent.**

I. Introduction

THE rapid advancement of medical technology, along with a growing focus on patient individuality make the traditional paternalistic decision-making model inadequate for modern healthcare. This is particularly evident in the treatment of major and chronic diseases, which usually require high levels of patient participation [1]. In response, Shared Decision-Making (SDM) emerges as a more inclusive healthcare model that actively involves patients and medical professionals in the development of treatment plans [2]. However, several factors hinder the implementation of SDM, including limited consultation time, poor communication skills among some healthcare providers, and patients' lack of confidence in their medical knowledge.

To overcome these barriers, efforts are made to automate the SDM process. A promising solution is the use of agent-based automated negotiation, which models the preferences and goals of both parties using utility functions and facilitates agreement through negotiation strategies. This technology shows success in various domains, offering efficient and effective decision-making results [3, 4, 5, 6].

While agent-based automated negotiation models show promise, many rely on linear utility functions, which are too simplistic to capture the complexity of real-world doctor-patient interactions [7, 8]. Medical decisions often involve nuanced subjective preferences that cannot be accurately represented by linear models. To address this limitation, fuzzy membership functions are introduced to represent the preferences of both doctors and patients, allowing for a more flexible and accurate modeling of their subjective preferences. This approach aligns more closely with the uncertainties and trade-offs inherent in medical decision-making, making it better suited for the complexities of SDM. For example, Lin et al. [9] proposes a model that frames preference negotiation in SDM as a problem of distributed fuzzy constraint satisfaction. This model involves designing doctor and patient agents to negotiate, with the results being transformed into treatment plans through a recommendation model.

Additionally, current SDM automated negotiation models often rely on fixed negotiation strategies based on machine learning techniques such as genetic algorithms and Bayesian algorithms[10, 11]. These fixed strategies lack adaptability to different types of users, such as patients and healthcare providers, as well as to their unique concession behaviors during negotiations. When the negotiation domain is wide, this results in a lower utility of the negotiation outcomes and necessitates additional rounds of negotiations to achieve consensus. Several studies propose adaptive strategies that utilize deep reinforcement learning algorithms to effectively address the limitations of fixed negotiation strategies. For example, Bakker et al. [12] propose the RLBOA negotiation model, which employs Q-learning within a reinforcement learning framework to enhance the adaptability of an agent's bidding strategy through experiential learning. However, this approach does not support continuous actions, limiting its applicability in more complex negotiation scenarios. Bagga et al. [7] introduce ANEGMA, a model that enables

Manuscript received October 9, 2024; revised March 21, 2025.

This work is supported by High Level Talent Project of Xiamen University of Technology under Grant (YSK20002R).

Sijie Wen is a postgraduate student from the Department of Computer Science and Technology of Xiamen University of Technology, Xiamen, 361024, China. (e-mail: 2222031156@stu.xmut.edu.cn).

Zhongqi Cai is a lecturer from the Department of Computer Science and Technology of Xiamen University of Technology, Xiamen, 361024, China. (e-mail: zhongqicai@xmut.edu.cn).

Chen Xin is a lecturer at the School of Data Science and Computer Science, Xiamen Institute of Technology, Xiamen, 361024, China. (e-mail: chenxin@xit.edu.cn).

Rende Hong is a postgraduate student from the Department of Computer Science and Technology of Xiamen University of Technology, Xiamen, 361024, China. (e-mail: 2222031189@stu.xmut.edu.cn).

Kaibiao Lin is a professor from the Department of Computer Science and Technology of Xiamen University of Technology, Xiamen, 361024, China. (e-mail: kblin@xmut.edu.cn).

Ping Lu is an associate professor from the School of Economic and Management of Xiamen University of Technology, Xiamen, 361024, China. (corresponding author to provide phone: 15060776129; e-mail: luping@xmut.edu.cn).

agents to learn negotiation strategies during concurrent bilateral negotiations in dynamic e-markets. By employing an Actor-Critic (AC) Network and utilizing model-free reinforcement learning, ANEGMA develops negotiation strategies pre-trained with synthetic market data, reducing exploration time in actual negotiations.

In summary, while each model contributes to the field, they predominantly focus on linear utility and preferences and rely on fixed negotiation strategies. This approach is insufficient to address the complexities of the SDM domain. Taking into account the above limitations and considering the preferences of both doctors and patients, this paper proposes an automated negotiation model for SDM based on fuzzy theory and the DRL algorithm for AC networks, named Actor-Critic (AC) Network-Based Agent Negotiation Model for SDM(ACANM-SDM). The model offers an innovative approach to resolving the optimization problem of offer and acceptance strategies, with the main contributions being the strategies. To overcome these limitations and better account for the nuanced preferences of both doctors and patients, this paper proposes a novel automated negotiation model for SDM, based on fuzzy theory and a DRL algorithm designed for AC networks, named ACANM-SDM. This model offers an innovative solution to optimize both offer and acceptance strategies, with the following key contributions:

(1) To facilitate automated negotiation and streamline the shared decision-making process, the model utilizes agent-based techniques to improve communication between doctors and patients. It incorporates trapezoidal fuzzy membership functions to represent the preferences of both parties, thereby addressing the limitations of linear utility models in complex SDM scenarios.

(2) The model creates a set of pre-trained data based on issue weights in doctor-patient negotiations to simulate users' concession behaviors. It employs two AC networks to develop strategies for bidding and acceptance, enabling the negotiation strategies to learn various concession patterns from different doctor-patient interactions. This approach effectively reduces negotiation time and enhances efficiency. Additionally, an alternating offer protocol is implemented to facilitate bilateral negotiation between the doctor and patient agents.

(3) Experimental results demonstrate that ACANM-SDM performs exceptionally well in SDM scenarios, providing robust decision support for both doctors and patients.

The rest of this paper is organized as follows: Section II reviews related work, offering an overview of existing methods in SDM and their limitations. Section III introduces the design and implementation of the ACANM-SDM model, including the negotiation environment, strategies, framework, and reward functions. Section IV details the experimental setup, results, and evaluations. Finally, Section V concludes the paper and suggests directions for future research.

II. Literature Review

This section provides a comprehensive review of the current state of research, focusing on three key aspects: the theoretical techniques of SDM, agent-based automated negotiation techniques, and the application of DRL in the field of automated negotiation.

A. The Conception and Related Theories of SDM

To advance the application of SDM, Charles et al. [13] define it as a decision-making model where both the doctor and the patient share information. They identify key features of SDM: the involvement of both parties (doctor and patient), mutual exchange of information, articulation of treatment preferences by both parties, and a final consensus on the selected treatment. However, communication skills, limited time, and patients' lack of medical knowledge significantly hinder their active participation in the decision-making process, thereby impacting the effective implementation of SDM.

To address these limitations, numerous studies propose approaches to improve the implementation of SDM. These approaches can be categorized into three types of models: (1) conceptual models, which define SDM processes and provide foundational guidance; (2) tools, such as Patient Decision Aids (PDAs), to facilitate practical implementation in clinical settings; and (3) measurement models, such as the Decision Conflict Scale (DCS) and the Shared Decision Making Questionnaire (SDM-Q), which assess the extent of patient involvement and uncertainty in decision-making.

Specifically, Elwyn et al. [14] propose a three-step SDM model in which the physician presents treatment options, discusses the details, and guides the patient in making a decision. However, this model does not fully emphasize the patient's role or account for post-decision steps. To address these gaps, Stiggelbout et al. [2] refine it with a four-step framework that supports patient preferences, ensures decisions are based on them, and includes post-decision plans. Both models serve as conceptual frameworks for SDM. To enhance practical implementation, tools like PDAs are introduced to improve patient access to information and communication [15], but current PDAs face challenges in addressing information asymmetry and fully meeting patient needs. Measurement tools, such as the Decision Conflict Scale and OPTION scale, assess SDM effectiveness but often rely on manual methods, leading to inefficiencies. Together, these tools aim to enhance patient involvement and decision-making in SDM.

While clinical analyses show that these tools improve understanding and communication [16, 17], they do not fully address the complex, dynamic nature of SDM [18]. Current research focuses on PDAs, clinical decision support systems, and physician communication skills training, but challenges remain in balancing patient preferences [19], ensuring medical reasonableness, and managing implementation costs [20].

To address these limitations, our study seeks to integrate SDM with Agent technology, delivering a comprehensive solution to enhance the effectiveness and practicality of SDM in clinical settings.

B. Agent-Based Automated Negotiation

Agent-based automated negotiation technology, which clarifies agents' preferences and goals using utility functions, significantly enhances the negotiation process in various applications. This methodology encourages structured communication and collaboration among agents, which in turn enables efficient conflict resolution and the achievement

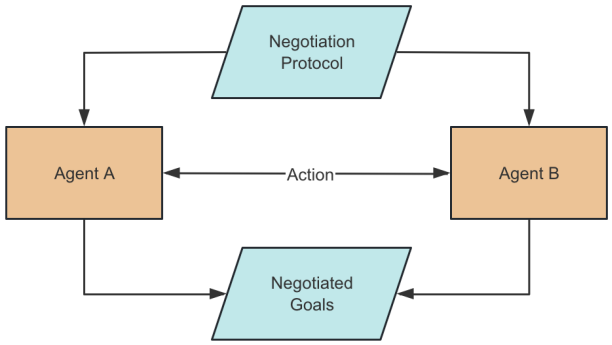

Fig. 1: Automated negotiation model.

of mutually beneficial outcomes. This technology is widely used in e-commerce [3, 21, 22], power trading [9, 23, 24], resource scheduling[4], control system[25], and workflow planning[5, 6], where this technology can optimize decision-making, reduce negotiation time, and improve efficiency, ultimately leading to more effective and equitable agreements in complex environments.

Specifically, an automated negotiation framework typically includes three key components: the negotiation agreement, negotiation goals, and negotiation agent, as illustrated in Figure 1. First, the agent identifies the negotiation goals, which determine what the parties want to achieve. Based on these objectives, the agent formulates offers and counter offers, which is presented through the negotiation agreement. As the negotiation unfolds, the agent continuously evaluates the responses and adjusts its strategy accordingly, ensuring that it is aligned with the negotiation objectives while adhering to the terms of the agreement. This iterative interaction between the agent and the agreement promotes a dynamic and efficient negotiation process.

The negotiation goal encompasses the issues that need resolution, which can be categorized into single-issue or multi-issue negotiations. Multi-issue negotiations are more complex because of the increased potential for conflicting preferences. Solutions, or offers, represent combinations of choices for each issue. These issues can be quantified (e.g., price, color, size). Linear issues involve straight forward metrics, while non-linear issues include more complex attributes.

The negotiation protocols outline the rules for agent interactions, with the Alternating Offers Protocol (AOP) being a common method [26]. In AOP, agents exchange offers turn by turn until time runs out or an offer is accepted, allowing them to make trade-offs and concessions to approach an optimal solution.

The design of an automated negotiation agent can be systematically analyzed through the components in the Bidding-Opponent-Accept (BOA) model proposed by Baarslag [27]. The BOA model negotiation Agent consists of three parts: (1) Bidding strategy: this part is used for the Agent to choose to make a counteroffer when it is dissatisfied with the opponent's offer; (2) Opponent model: when the agent faces an adversary with incomplete preference information, the adversary model predicts the future actions or complete preferences of the adversary by its behaviors and history of offers; and (3) Acceptance strategy: this component guides the Agent on the circumstances in which to accept the adversary's offer. For example, Elghamrawy [28] develops a multi-intelligence negotiation model for doctors, which uses an adaptive bidding strategy to improve diagnosis efficiency and reduce conflict resolution time. Sun et al. [29] proposes a bilateral price negotiation strategy based on Bayesian classification (for opponent modeling) and Q-learning (for generating a counter-offer).

Despite these advancements, existing models often require predefined strategies and struggle with dynamic preferences. Our research addresses this by incorporating DRL algorithms with agent technology, enabling models to autonomously adapt and learn. This approach transforms traditional SDM models into automated frameworks, improving efficiency and decision quality by reducing human intervention and accounting for dynamic preferences. We are refining these strategies using AC network DRL to optimize negotiation outcomes further [10, 11].

C. Deep Reinforcement Learning in Automated Negotiation

Previous automated negotiation models typically utilize machine learning to develop fixed negotiation strategies[30]. However, these fixed strategies often lack adaptability to different user types, such as patients and healthcare providers, along with their distinct concession habits. This limitation can result in reduced personal utility and social welfare for both parties involved in the negotiation. Consequently, recent studies implement DRL to better understand user behavior and effectively address the conflicts of preferences between the two parties[7, 8].

Deep reinforcement learning is widely used in many fields and has achieved good results[31, 32, 33]. DRL enhances automated negotiation in several ways. It dynamically learns adaptive strategies through real-time interactions, optimizes decision-making based on user preferences, and effectively manages complex multi-party interactions. These improvements lead to greater overall satisfaction and increased social welfare. Recent advancements in DRL have significantly enhanced the performance of automated negotiation systems. Bagga et al. [7] propose ANEGMA, a model that allows agents to learn negotiation strategies during synchronized bilateral negotiation in dynamic e-markets. Employing an AC network develops negotiation strategies with model-free reinforcement learning, pre-trained with synthetic market data to reduce exploration time in actual negotiations. Chang [8] investigates DRL in negotiation, focusing on agents' abilities to exploit, adapt, and cooperate. Two AC networks are trained for bidding and acceptance strategies against time-based and behavior-based agents, and also through self-play. The results demonstrate the effectiveness of the Cauchy distribution for sampling offers but highlight limitations in issue scalability and handling nonlinear problems.

III. Method

Negotiation protocols outline the rules and frameworks for agent interactions, with the Alternating Offers Protocol being a common method [26]. In the Alternating Offers Protocol, agents exchange offers turn by turn until either the time limit is reached or an offer is accepted. This process enables them

to make trade-offs and concessions, gradually moving toward an optimal solution.

This section details the design and implementation of the ACANM-SDM model proposed in this study. First, we set up the consultation environment and negotiation protocols, and specify the roles and behaviors of the Doctor Agent (DA) and the Patient Agent (PA) in the negotiation process. Next, we set a theoretical time limit for negotiation to simulate real-world decision pressure, and introduce the concept of time sensitivity to reflect the effect of time on the utility of negotiation. On this basis, the negotiation strategies in the ACANM-SDM model, including time-dependent and behavior-dependent strategies, and their applicability under incomplete information are further elaborated. Finally, the framework of the ACANM-SDM model is described in detail, including the algorithms and pre-training process for generating and training data, the architectural design of the Offer Net and Accept Net, and the construction of the reward function. Together, these components form a complete DRL system for optimizing the negotiation process between doctors and patients.

A. Negotiation Settings

Suppose that the negotiation environment E consists of a DA and a PA, which negotiate with each other over an issue domain, which consists of n different independent issues $D(I_1, I_2, ...I_n)$, each issue is taken to consist of a finite set of k possible discrete or continuous values, and the discrete values are considered in the experiment.

Before agents start negotiating and exchanging offers, they must agree on a negotiation protocol P that determines the effective moves agents can make in any negotiation state. In this paper, we use an alternating offer protocol in which the DA starts the negotiation by sending an offer to the PA, which the PA can either accept or reject; if it accepts, the negotiation ends with an agreement, otherwise, the PA makes a counter-offer to the DA. This process of sending offers continues until one of the agents accepts the offer (i.e., successful negotiation) or reaches a deadline (i.e., negotiation fails).

For theoretical and practical reasons, we impose real-time deadlines on the negotiation process. The pragmatic reason is that without deadlines, the negotiation may go on forever, especially if there is no discount factor. Secondly, in the case of unconstrained time, the intelligentsia can simply try out a large number of proposals to understand the preferences of the opponent. However, considering real-time deadlines poses many challenges, e.g., agents should be more willing to concede near the deadline because interruptions have zero utility (or retained utility, if any) for both agents; real-time deadlines also make it necessary to employ a strategy for deciding when to accept an offer; and deciding when to accept requires some prediction about whether a significantly better opportunity will arise in the future.

In addition, we assume that negotiations are time-sensitive, i.e., time affects the utility of both parties to the negotiation. In more detail, the value of an offer decreases over time.

B. Negotiation Strategy

With perfect information, it is possible to determine the optimal bidding strategy [34], however perfect information is unlikely to be available in bargaining as agents are reluctant to give up their preferences for fear of being exploited. This promotes the development of negotiation strategies under imperfect information. These strategies are broadly classified as either time-dependent or behavior-dependent based on the decision function that maps offer states to the target utility [35].

1) Time-Dependent Strategies: Time-dependent strategies involve decision-making processes that respond dynamically to temporal changes, such as the timing of actions or the duration of negotiations. These strategies recognize that the optimal course of action may vary over time, often incorporating elements like deadlines or time-based incentives to influence behavior. In each round, agents compute their decision utility to decide whether they accept the offer or not. The utility is computed by Equation (1).

$$u(t) = P_{min} + (P_{max} - P_{min}) * (1 - F(t)), \quad (1)$$

P_{max} and $P_{\text{min}} \in [0, 1]$, thus parameterizing the range of offers, and typically, $F(t)$ is parameterized as an exponential function, as shown in Equation (2).

$$F(t) = k + (1 - k) * (\frac{t}{T})^{\frac{1}{c}}, \quad (2)$$

where c is the concession factor, and k is typically set to 0 for simplicity. When $0 < c < 1$, the agent concedes gradually towards the end of the negotiation and is referred to as a Boulware agent. If $c > 1$, the agent concedes quickly, offering its reservation value, thus earning the label of a Conceder. Specifically, when $c = 1$, the agent's decision utility decreases linearly [36].

2) Behavior-Dependent Strategies: Behavior-dependent strategies, focus on adapting actions based on the observed behaviors of other agents or participants. The best known is Tit-For-Tat (TFT), whereby cooperation is generated through reciprocity. Its three core principles are: never be the first to defect; retaliate if provoked; and forgive after retaliating. The relative TFT strategy reciprocates by offering concessions that are proportional to the concessions made by the opponent in the previous δ rounds, as described in the process shown in Equation (3).

$$x_{a\to b}^{t_{n+1}}[j] = min(max(\frac{x_{a\to b}^{t_{n-2\sigma}}[j]}{x_{b\to a}^{t_{n-2\sigma+2}}[j]} x_{a\to b}^{t_{n-1}}[j], min_j^a), max_j^a), \quad (3)$$

here, $x_{a\to b}^{t_{n+1}}[j]$ denotes the action taken by agent a towards agent b at time step $n+1$. The term $x_{a\to b}^{t_{n-2\sigma}}[j]$ reflects the action from agent a to agent b at an earlier time step, specifically $n-2\sigma$. Similarly, $x_{b\to a}^{t_{n-2\sigma+2}}[j]$ indicates the action of agent b towards agent a at time $n-2\sigma+2$. The variable $x_{a\to b}^{t_{n-1}}[j]$ captures the action from agent a to agent b at time $n-1$. Lastly, min_j^a and max_j^a represent the minimum and maximum bounds for agent a's actions, ensuring that the resulting strategy stays within predefined limits. This formula promotes adaptive behavior by enabling agent a to adjust its actions based on the historical behaviors of both itself and agent b.

C. ACANM-SDM Automated Negotiation Modelling

This paper designs an SDM negotiation model based on fuzzy theory and AC network DRL algorithm, which contains

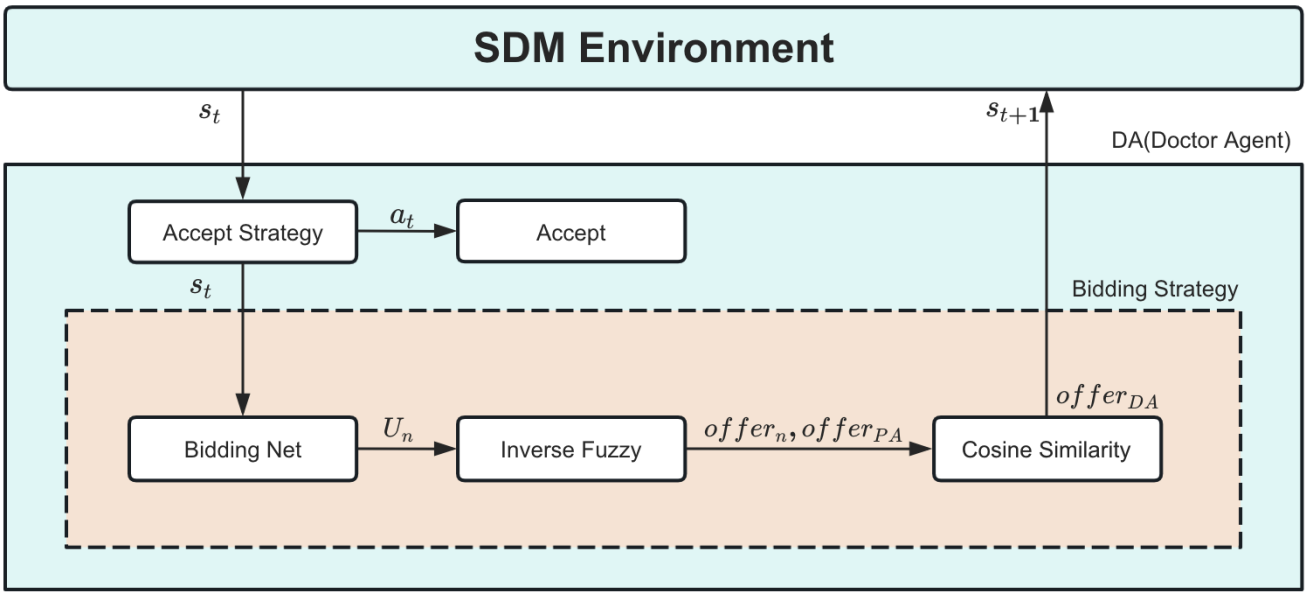

Fig. 2: ACANM-SDM model diagram.

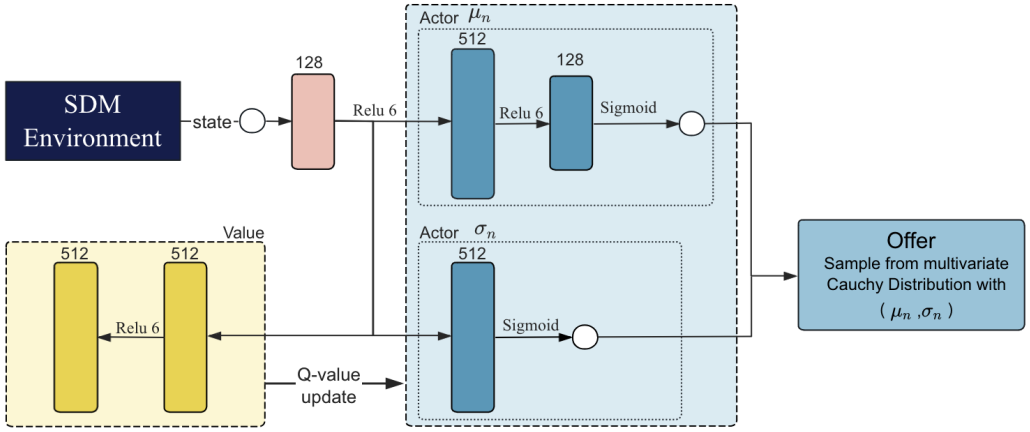

Fig. 3: Offer net architecture.

five modules as shown in Fig. 2. Module 1 is the SDM environment module, which takes one from the optimal offer set as the current round of bids and sends a state s_t to the opponent, which includes the current round of offer and the current time step. Module 2 is the acceptance strategy module, which inputs the state s_t into the module to determine whether to accept the current bid. If accepted, the negotiation ends. If not accepted, s_t is input into the bidding strategy module. Module 3 is the bidding strategy network module, which calculates the utility of the counter-offer of all issues. Module 4 is the anti-fuzzy membership function module, which inputs the calculated utility (U_n) set into the anti-fuzzy membership function to obtain all issue values corresponding to the utility. Module 5 Similarity calculation module, input all issue values $offer_n$ and the opponent's offer in-state s_t into this module, get the offer with the highest similarity to the opponent, and send the state to the opponent. A key point of interest is that the acceptance and bidding strategies need to be trained, and in the initial state, the network acts as a stochastic strategy. We pre-train the two-strategy model using real SDM issue weights combined with a time-dependent strategy, greatly reducing resource and time overheads during the exploration period.

D. Pre-training

As deep learning was integrated into the reinforcement learning algorithm, we were able to pre-train the actor network module using supervised learning. In this way, the exploration cost of the training process can be greatly reduced, and the 1st thing we need to address in order to pre-train the strategy is to collect suitable actor network labeled data. We generate 20 sets of training data using Equation (4) and then pre-train the actor network using backpropagation. After pre-training the actor network, the pre-training of the Critic network is continued by fixing the form of the actor network.

$$U_{issue}(u(t), V_{issue}) = \frac{u(t)}{1 - V_{issue}}, \tag{4}$$

Equation (4) generates 20 training datasets, where V_{issue} represents the weights assigned by doctors and patients to different issues, and $u(t)$ uses the utility in equation (1).

E. Offer Net Architecture

If the agent rejected the offer, the agent would now take the same input and decide to counteroffer. Since we have five issues, the offer network outputs a vector $o = [0, 1]^5$ representing the utility of the five issues in the offer, and DRL is achieved by continuous control, sampling from the multivariate Beta distribution. The Beta distribution is defined on the interval [0, 1] and is defined by two shaping parameters α and β, the process is shown in Equation (5).

$$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} with B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \quad (5)$$

The bidding strategy model of the doctor-patient Agent is shown in Fig. 3, and the input current state s_t is still a six-element vector containing the opponent's offer (cost, risk, treatment period, effectiveness, convenience) and the current time. It is first determined whether the agent accepts the current offer, acceptance ends the negotiation, and rejection of the current offer results in a counter-offer. The base layer outputs to the value network and ten other neural blocks are used to estimate the distributional parameters, the five means (μ), and variances (σ). These six variables are then used to sample the offers with a Beta distribution while this sample is used to calculate the loss. The loss calculation for the bidding strategy network is shown in Equations (6)-(9).

$$TD_{\text{Loss}} = \text{Reward} - q(s_t, a_t, \vec{\omega_t}), \quad (6)$$

Equation (6) is used to calculate the Temporal Difference loss, representing the difference between the current reward and the estimated Q-value, denoted as $q(s_t, a_t, \vec{\omega_t})$, which indicates the expected return for state s_t and action a_t. Equation (7) is a square loss function, which is used to calculate the square of TD_{Loss}.

$$C_{loss} = (TD_{Loss})^2, \quad (7)$$

this helps optimize the network to better align with the value estimates.

Equation (8) denotes the logarithm of the probability of taking action under policy π.

$$\text{LogProbs} = \ln P(a \mid \pi), \quad (8)$$

here, $P(a \mid \pi)$ is the probability of action of a given policy π.

Equation (9) is used to calculate Actor Loss, this combines the log probability and an entropy term to optimize the actor network. The loss is the product of the TD error and log probability, with a negative sign to maximize expected returns. The entropy term encourages exploration, preventing the policy from becoming too deterministic.

$$A_{\text{loss}} = -(\text{LogProbs} + \text{Entropy}) \cdot TD_{\text{Loss}}. \quad (9)$$

F. Accept Net Architecture

The acceptance strategy model is shown in Fig. 4. The input s_t is a six-element vector containing the adversary's offer (cost, effective, side effects, risk, convenience) and the current time. At each time step, the acceptance strategy network accepts the adversary's offer, encoding it into 512 hidden states using two affine Relu 6 pairs. This base layer is shared between the agent and the value network. The actor accepts the embedded state and outputs two logit values, which are used to select the appropriate action (0: reject, 1: accept), and the value net outputs an estimate of the expected return. Then, the critic loss is computed by obtaining the mean square error (MSE) of the time difference observed reward minus the forward pass of the value net, the loss is computed as shown in Equations (6)-(8), and Equations (10).

The calculation method of actor loss (Aloss) is as shown in Equation (10).

$$A_{loss} = -LogProbs * TD_{Loss}, \quad (10)$$

this loss is associated with optimizing action selection and determining how to adjust the policy to achieve higher rewards.

G. Reward Function

For the setting of the reward function, the acceptance strategy network and the bidding strategy network share the same reward function. With deadline T, issue weight ω, and issue utility of final offer x, the rewards of doctor and patient Agents are shown in Equation (11).

$$R_t = \begin{cases} \omega^t \cdot x & \text{if } t_f < T \text{ and } t_f \neq 0, \\ 0 & \text{if } t_f = T(conflictdeal). \end{cases} \quad (11)$$

In summary, our reward function is designed to encourage timely and successful negotiations by providing a positive reward value when the agent concludes a deal.

IV. Experiment

In this section, we first define evaluation metrics in Section IV-A, then design the experimental to evaluate ACANM-SDM in Section IV-B. Finally, we present the results and analysis in Section IV-C.

A. Evaluation Metrics

In this work, we adopt four evaluation metrics:

- Negotiation Rounds ($Round$): $Round$ is the number of rounds required for a successful negotiation, which measures the speed of convergence of the negotiation model, with faster convergence indicating less influence by time constraints.
- Doctor Utility (U_{DA}): U_{DA} is the doctor's satisfaction with the negotiation result.
- Patient Utility (U_{PA}): U_{PA} is the patient's satisfaction with the negotiation result.
- Social Welfare Utility (U_{soc}): U_{soc} is the sum of physician utility and patient utility, and a larger value indicates that the negotiation model can obtain more benefits for the decision-maker.
- Utility difference(Difference): Utility difference is the difference in utility between the two parties for the negotiation results, reflecting whether the model can balance the interests of all decision-makers. The smaller the difference, the fairer the negotiation model.

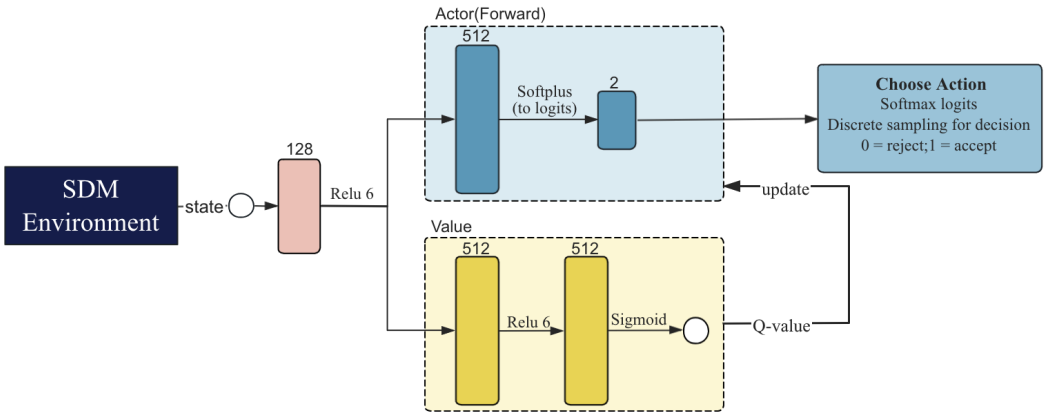

Fig. 4: Accept net architecture.

TABLE I: Doctor and patient preferences

Issue	PA preference		DA preference	
	Evaluation Functions	Weights	Evaluation Functions	Weights
cost	(0,1,3.5,4.5)	$(0.3)_W$	(3.5,4,8,8)	$(0.15)_W$
effective	(7,9,10,10)	$(0.25)_W$	(5,7,8,10)	$(0.3)_W$
side effects	(0,0,5,10)	$(0.2)_W$	(1,10,15,25)	$(0.25)_W$
risk	(0,0,5,10)	$(0.15)_W$	(5,10,15,25)	$(0.2)_W$
convenience	(7,9,10,10)	$(0.1)_W$	(6,7,8,10)	$(0.1)_W$

TABLE II: Negotiation domain

Issue	**Domain**
cost	$0-8,000$
effective	Levels 1-10
side effects	$0-100\%$
risk	$0-100\%$
convenience	Levels 1-10

B. Experimental Design

We design two experiments to verify the negotiation efficiency of ACANM-SDM and evaluate the ACANM-SDM effectiveness in the three different sizes of issue domains. The experimental design details are as follows:

Experiment 1: To evaluate the performance of the model, this experiment compares the negotiation performance of the ACANM-SDM model with other negotiation models in a negotiation environment with a different number of round of deadlines. The comparison models include the Boulware, Linear, and Conceder models with time-dependent strategies [36], and a fuzzy constraint-based SDM automated negotiation model, FCAN [9]. To simulate the negotiation environment under different time constraints, five cut-off rounds (Deadlines) are set, which are 10, 15, 20, 25, and 30 rounds, and the number of issues is fixed at five.

Experiment 2: To verify the performance of the ACANM-SDM model in a negotiation environment with a different number of issues and to compare the negotiation performance of the ACANM-SDM model with that of other negotiation models, the number of issues in this experimental design is set at 3, 5, and 7, and the number of negotiation deadline rounds is fixed at 20.

Experiment 3: To prove that the model can be effective in negotiations for different doctors and patients and to obtain negotiation results with high satisfaction for both parties, the model uses three different doctor-patient preference data as input for comparison, the number of negotiation deadline rounds is fixed to 15.

To ensure the reliability of the negotiation results, each negotiation will be simulated 50 times in environments with different time constraints or number of issues, respectively. All experiments use the same SDM negotiation domain information collected by Lin et al. [9] in real healthcare environments, with the PA and DA preference data shown in Table I and the negotiation domain configurations shown in Table II. The utility preferences consist of a set of trapezoidal fuzzy affiliation functions. The set of issue weight preferences W is composed of decimals from 0 to 1 and the sum of all issue weights is 1.

To ensure the reliability of the negotiation results, each negotiation will be simulated 50 times in environments with different time constraints or number of issues, respectively. All experiments use the same SDM negotiation domain information collected by Lin et al. [9] in real healthcare environments, with the PA and DA preference data shown in Table I and the negotiation domain configurations shown in Table II. The utility preferences consist of a set of trapezoidal fuzzy affiliation functions. The set of issue weight preferences W is composed of decimals from 0 to 1 and the sum of all issue weights is 1.

C. Experimental Results

In this section, we assess the effectiveness of ACANM-SDM by evaluating the *Round*, U_{DA}, U_{PA}, and U_{soc} through **Experiment 1**, **Experiment 2** and **Experiment 3**.

1) Comparative Experiments with Different Number of Deadline Rounds: Building on the design of Experiment 1, this experiment compares the ACANM-SDM model with the Boulware, Linear, Conceder, and FCAN models. The experimental results demonstrate that ACANM-SDM outperforms other models in U_{soc}, U_{DA} and Round, U_{PA} second only to the FCAN model. The specific experimental results are shown in Fig. 5, Fig. 6, and Fig. 7, as well as Table III, in which the bold part of the table marks the best performance in each time-constrained environment. More specifically, Fig. 5 illustrates the social welfare utility results for all models in the consultation environment with different cut-off rounds, and the ACANM-SDM model outperforms the other comparative models for all cut-off rounds, with improvement effects ranging from 4% to 25%. In terms of personal utility, the experimental results are shown in Fig. 6 and Fig. 7, the results of all the models are more variable, the doctor's utility of the ACANM-SDM model is higher than the other models. The patient's utility of the FCAN model is higher than the different models. Still, the difference between the doctor's and the patient's utility of the ACANM-SDM model is smaller than that of the FCAN model, which proves that the fairness of the ACANM-SDM model is better.

In terms of the number of negotiation rounds, the experimental results are shown in Fig. 8, the number of negotiation rounds of Linear, Conceder, and Boulare models increases with the increase of the number of cut-off rounds, the negotiation discourse of FCAN can be stable at around 9 rounds in the negotiation environment with different cut-off rounds, and the ACANM-SDM model controls the number of negotiation rounds at around 4 and can remain stable across different deadline rounds.

Combining the experimental results for the *Round*, U_{DA}, U_{PA}, and U_{soc}, the ACANM-SDM model possesses a better performance in terms of the number of consultation rounds, physician utility, and social welfare utility, and the obtained patient utility is lower than that of the FCAN model but still higher than that of the other comparison models. The above experimental results show that the proposed model in this study can achieve higher social welfare utility and fairness with fewer consultation rounds in a variety of different cut-off round environments. This implies that the proposed model in this study is more adaptable in the face of the complex and changing SDM real-world consultation environments, especially in environments where the general consultation time is tight.

2) Comparative Experiments with Different Number of Issues: The experiment distinguishes the negotiation environments set up with different numbers of issues into small negotiation domains (i.e., the number of issues is 3 issues), medium-sized negotiation domains (5 issues), and large negotiation domains (7 issues), and introduces the model mentioned in Experiment 1 for comparison. The specific consultation results are shown in Fig. 9, Fig. 10 and Table IV.

ACANM-SDM outperforms the other models in both

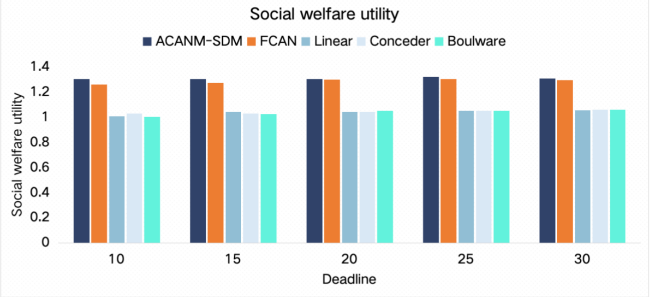

Fig. 5: Comparison of ACANM-SDM VS other agents in terms of social welfare utility

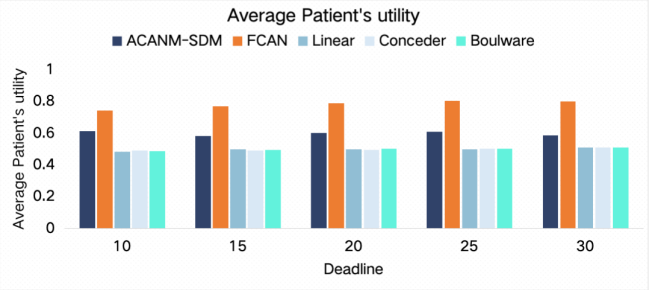

Fig. 6: Comparison of ACANM-SDM VS other agents in terms of patient agent utility

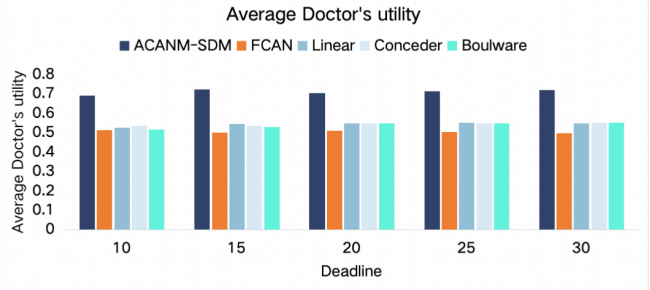

Fig. 7: Comparison of ACANM-SDM VS other agents in terms of doctor agent utility

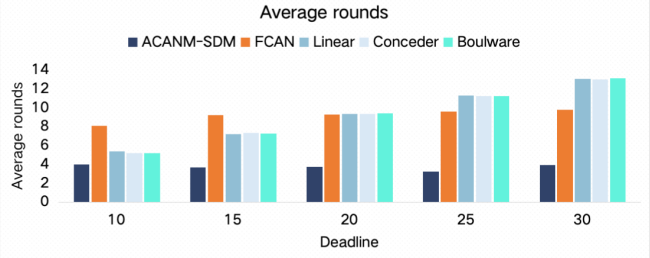

Fig. 8: Comparison of ACANM-SDM VS other agents in terms of average rounds

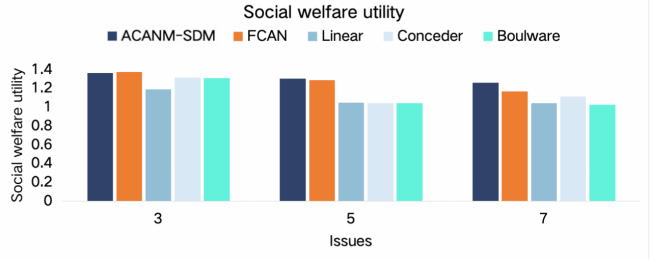

Fig. 9: Comparison of ACANM-SDM VS other agents in terms of social welfare utility

TABLE III: Experimental results for different numbers of deadline rounds

Deadline	Evaluation indicators	ACANM-SDM	FCAN	Boulware	Linear	Conceder
10	U_{soc}	**1.307**	1.259	1.006	1.009	1.029
	U_{PA}	0.614	**0.745**	0.488	0.482	0.491
	U_{DA}	**0.693**	0.514	0.518	0.528	0.537
	Round	**4.000**	8.090	5.200	5.360	5.200
15	U_{soc}	**1.306**	1.274	1.027	1.044	1.030
	U_{PA}	0.582	**0.772**	0.497	0.498	0.493
	U_{DA}	**0.724**	0.502	0.531	0.547	0.537
	Round	**3.660**	9.210	7.300	7.200	7.360
20	U_{soc}	**1.306**	1.299	1.051	1.045	1.045
	U_{PA}	0.601	**0.789**	0.501	0.498	0.497
	U_{DA}	**0.705**	0.511	0.550	0.548	0.548
	Round	**3.740**	9.260	9.440	9.380	9.360
25	U_{soc}	**1.323**	1.307	1.052	1.053	1.052
	U_{PA}	0.608	**0.804**	0.503	0.501	0.502
	U_{DA}	**0.715**	0.503	0.548	0.552	0.550
	Round	**3.220**	9.580	11.240	11.300	11.240
30	U_{soc}	**1.308**	1.298	1.063	1.056	1.063
	U_{PA}	0.587	**0.799**	0.512	0.509	0.512
	U_{DA}	**0.721**	0.498	0.551	0.548	0.551
	Round	**3.960**	9.770	13.900	13.080	12.980

TABLE IV: Experimental results at different number of issues

Issue Number	Evaluation indicators	ACANM-SDM	FCAN	Linear	Conceder	Boulware
3	U_{soc}	1.365	**1.378**	1.190	1.318	1.312
	Round	**3.720**	9.000	8.000	6.000	10.000
5	U_{soc}	**1.306**	1.288	1.051	1.045	1.045
	Round	**3.740**	9.160	9.440	9.380	9.360
7	U_{soc}	**1.264**	1.169	1.043	1.114	1.026
	Round	**5.450**	9.940	11.440	12.030	10.982

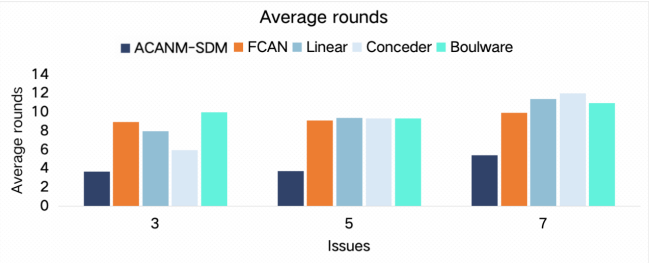

Fig. 10: Comparison of ACANM-SDM VS other agents in terms of average rounds

medium and large consultation domains, achieving higher social welfare and fewer consultation rounds. This suggests that ACANM-SDM possesses a stronger spatial search capability than ANFGA and is adapted to SDM environments with a variety of number of issue domains and complex treatment options. In the small consultation domain, ACANM-SDM outperforms the time-dependent strategy family model in terms of social welfare but is second to the FCAN model by about 1%. This is because ACANM-SDM takes time urgency into account and aims to achieve optimal individual utility and social welfare for doctors and patients with fewer consultation rounds.

3) Comparative experiment under different doctor-patient preferences: To demonstrate that the model can conduct effective negotiations for different doctors and patients and obtain high satisfaction negotiation results for both parties, the model uses three different doctor-patient preference data as input for comparison. The data sets are shown in Fig. 11 to Fig. 15 . This comparative experiment can prove that the model can achieve good negotiation results for ordinary doctors and patients, both in cases of large conflicts and in cases of small conflicts. The specific results of the experiment are shown in Tables 5 and 6. The experimental results show that when the conflict between

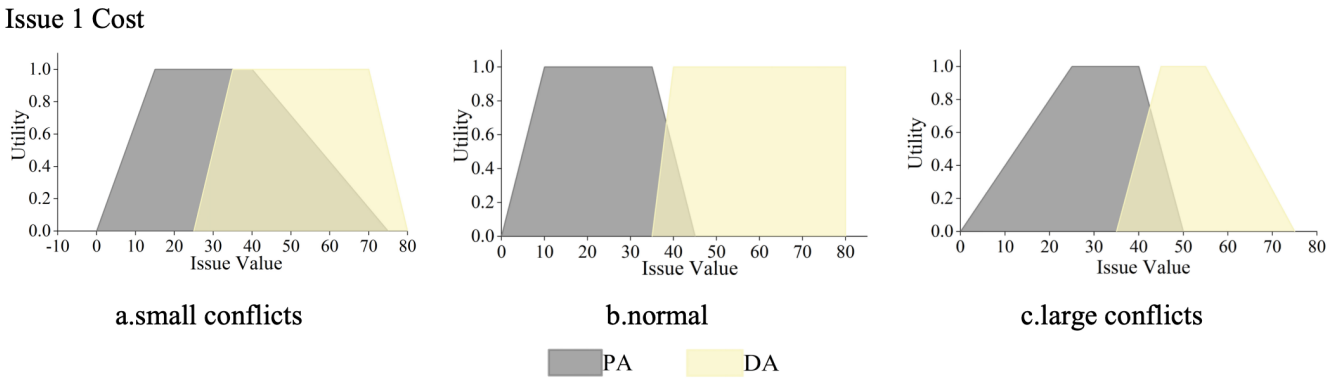

Fig. 11: Three different groups of doctor-patient preferences in issue 1

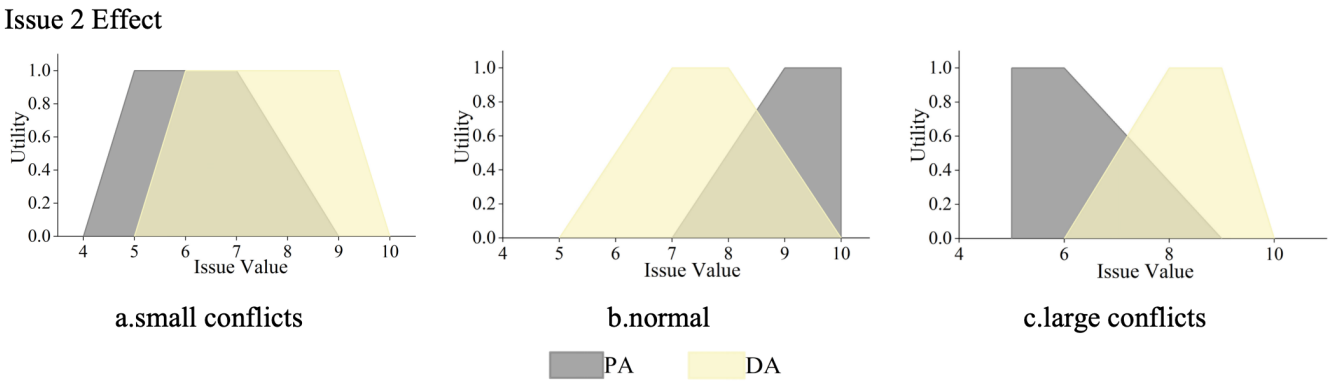

Fig. 12: Three different groups of doctor-patient preferences in issue 2

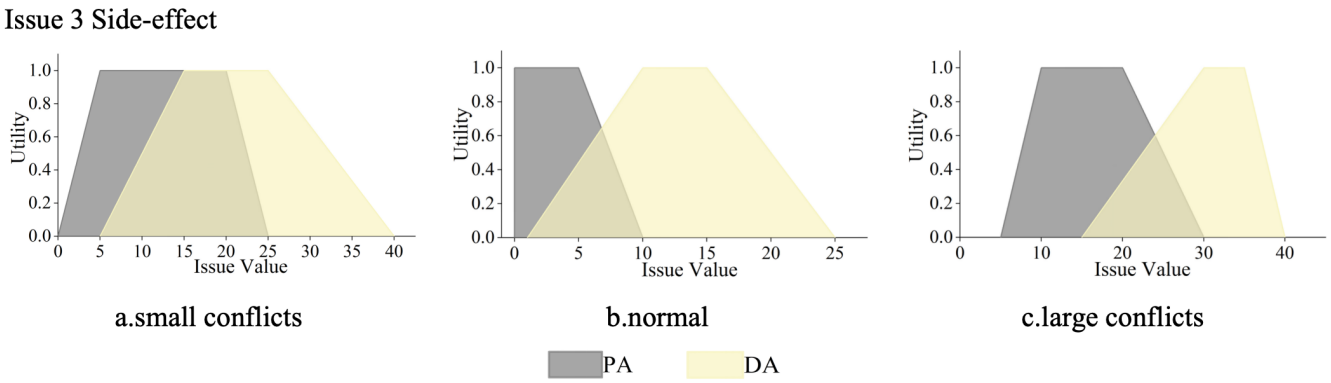

Fig. 13: Three different groups of doctor-patient preferences in issue 3

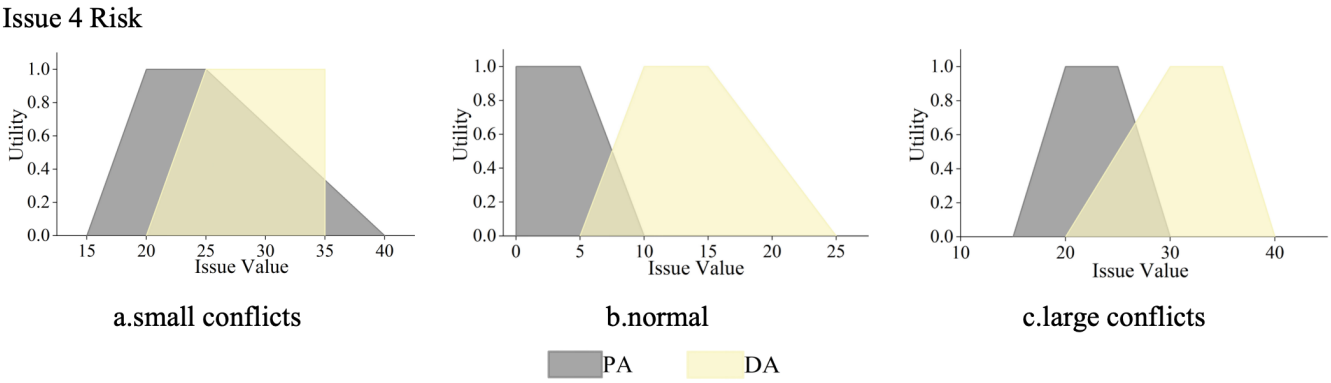

Fig. 14: Three different groups of doctor-patient preferences in issue 4

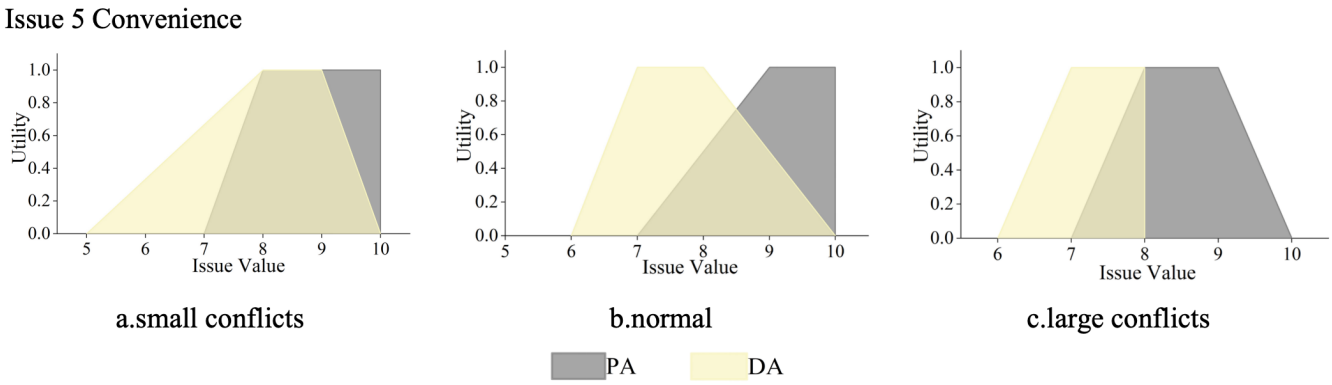

Fig. 15: Three different groups of doctor-patient preferences in issue 5

TABLE V: Performance Comparison of Different Models with High Doctor-Patient Preference Conflict

Evaluation Metrics	ACANM-SDM	FCAN	Boulware	Linear	Conceder
U_{soc}	**1.196**	1.098	1.159	1.131	1.089
U_{PA}	**0.572**	0.531	0.550	0.540	0.527
U_{DA}	**0.624**	0.567	0.609	0.591	0.561
Fairness	0.052	0.036	0.059	0.051	0.034
Round	**2.78**	8	6.74	8.16	14.14

TABLE VI: Performance Comparison of Different Models with High Doctor-Patient Preference Conflict

Evaluation Metrics	ACANM-SDM	FCAN	Boulware	Linear	Conceder
U_{soc}	**1.797**	1.783	1.729	1.749	1.830
U_{PA}	**0.90**	0.901	0.825	0.852	0.930
U_{DA}	**0.897**	0.882	0.903	0.897	0.899
Fairness	**0.003**	0.019	0.078	0.045	0.031
Round	**2.2**	4.4	2.62	3.46	8.84

doctor-patient preferences is large, both the social welfare and the personal utility of both doctors and patients decrease. However, the ACAN model shows significant advantages compared to the baseline models. Specifically, the ACAN model improved the optimal result of the baseline model by 0.037 in the comparison experiment, and improved by 0.022, 0.032, and 0.045 respectively; the utility of both doctors and patients was improved by 0.015, 0.033, and 0.063 respectively; the number of negotiation rounds was reduced by 5.22, 3.96, 5.38, and 11.36. The experimental results show that even when there is a large conflict between doctor-patient preferences, the ACAN model can still achieve an ideal negotiation effect within fewer rounds of negotiation and generate a quote that can make doctors and patients more satisfied. The experimental results show that when the conflict between doctor-patient preferences is small, the four models all achieve good negotiation results, among which the Conceder model has the highest effect and the ACANM-SDM is the second best. However, ACANM-SDM requires the smallest average number of negotiation rounds, which can save the time cost of diagnosis and treatment in a tense doctor-patient environment.

V. Conclusion and Future Work

In this paper, we propose an automated negotiation model based on deep reinforcement learning (ACANM-SDM) that simulates doctor-patient communication and optimizes negotiation strategies through an intelligent agent. Specifically, ACANM-SDM first incorporates an AC network and fuzzy theory to reduce exploration costs via pre-training. Then it uses fuzzy membership functions to realistically represent doctor-patient preferences. Experimental results show that ACANM-SDM improves social welfare by 26% and reduces the average number of negotiation rounds by 50% compared to traditional negotiation models, thus validating its effectiveness in complex SDM scenarios. However, despite ACANM-SDM's superior performance in experimental settings, its generalizability to diverse healthcare contexts and cultural settings requires further enhancement. Therefore, our future work will focus on enhancing the model's adaptability across various clinical environments and validating its efficacy in real-world medical practice.

References

[1] D. Stacey, F. Légaré, K. Lewis, M. J. Barry, C. L. Bennett, K. B. Eden, M. Holmes-Rovner, H. Llewellyn-Thomas, A. Lyddiatt, R. Thomson *et al.*, "Decision aids for people

facing health treatment or screening decisions," *Cochrane database of systematic reviews*, no. 4, 2017.

[2] A. M. Stiggelbout, A. H. Pieterse, and J. C. De Haes, "Shared decision making: concepts, evidence, and practice," *Patient education and counseling*, vol. 98, no. 10, pp. 1172–1179, 2015.

[3] P. Bagga, N. Paoletti, B. Alrayes, and K. Stathis, "ANEGMA: an automated negotiation model for e-markets," *Autonomous Agents and Multi-Agent Systems*, vol. 35, no. 2, p. 27, 2021.

[4] A. Almutawakel, O. Kazar, M. Bali, H. Belouaar, and A. Barkat, "Smart and fuzzy approach based on CSP for cloud resources allocation," *International Journal of Computers and Applications*, vol. 44, no. 2, pp. 117–129, 2022.

[5] C.-Y. Hsu, B.-R. Kao, K. R. Lai *et al.*, "Agent-based fuzzy constraint-directed negotiation mechanism for distributed job shop scheduling," *Engineering Applications of Artificial Intelligence*, vol. 53, pp. 140–154, 2016.

[6] Z. Zahedi, S. Sengupta, and S. Kambhampati, "'Why didn't you allocate this task to them?'Negotiation-Aware Task Allocation and Contrastive Explanation Generation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 9, 2024, pp. 10 243–10 251.

[7] P. Bagga, N. Paoletti, B. Alrayes, and K. Stathis, "A deep reinforcement learning approach to concurrent bilateral negotiation," *arXiv preprint arXiv:2001.11785*, 2020.

[8] H.-C. H. Chang, "Multi-issue negotiation with deep reinforcement learning," *Knowledge-Based Systems*, vol. 211, p. 106544, 2021.

[9] K. Lin, Y. Liu, P. Lu, Y. Yang, H. Fan, and F. Hong, "Fuzzy constraint-based agent negotiation framework for doctor-patient shared decision-making," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, p. 218, 2022.

[10] K.-B. Lin, Y. Wei, Y. Liu, F.-P. Hong, Y.-M. Yang, and P. Lu, "An opponent model for agent-based shared decision-making via a genetic algorithm," *Frontiers in Psychology*, vol. 14, p. 1124734, 2023.

[11] X. Chen, P. Lu, K.-B. Lin, and J.-Y. Zeng, "Particle Swarm Optimization-based Agent Negotiation Framework to Support Shared Decision-Making," vol. 8, no. 3, pp. 758–775, 2023.

[12] J. Bakker, A. Hammond, D. Bloembergen, and T. Baarslag, "RLBOA: A Modular Reinforcement Learning Framework for Autonomous Negotiating Agents." in *AAMAS*, 2019, pp. 260–268.

[13] C. Charles, A. Gafni, and T. Whelan, "Shared decision-making in the medical encounter: what does it mean?(or it takes at least two to tango)," *Social science & medicine*, vol. 44, no. 5, pp. 681–692, 1997.

[14] G. Elwyn, A. Edwards, M. Wensing, K. Hood, C. Atwell, and R. Grol, "Shared decision making: developing the OPTION scale for measuring patient involvement," *BMJ Quality & Safety*, vol. 12, no. 2, pp. 93–99, 2003.

[15] J. F. Bridges, E. W. de Bekker-Grob, B. Hauber, S. Heidenreich, E. Janssen, A. Bast, J. Hanmer, A. Danyliv, E. Low, J. C. Bouvy *et al.*, "A roadmap for increasing the usefulness and impact of patient-preference studies in decision making in health: a good practices report of an ISPOR task force," *Value in Health*, vol. 26, no. 2, pp. 153–162, 2023.

[16] M. Slade, "Implementing shared decision making in routine mental health care," *World psychiatry*, vol. 16, no. 2, pp. 146–153, 2017.

[17] T. Agoritsas, A. F. Heen, L. Brandt, P. Alonso-Coello, A. Kristiansen, E. A. Akl, I. Neumann, K. A. Tikkinen, T. Van der Weijden, G. Elwyn *et al.*, "Decision aids that really promote shared decision making: the pace quickens," *Bmj*, vol. 350, 2015.

[18] N. Joseph-Williams, D. Williams, F. Wood, A. Lloyd, K. Brain, N. Thomas, A. Prichard, A. Goodland, H. McGarrigle, H. Sweetland *et al.*, "A descriptive model of shared decision making derived from routine implementation in clinical practice ('Implement-SDM'): Qualitative study," *Patient Education and Counseling*, vol. 102, no. 10, pp. 1774–1785, 2019.

[19] B. Büdenbender, A. K. Köther, B. Grüne, M. S. Michel, M. C. Kriegmair, and G. W. Alpers, "When attitudes and beliefs get in the way of shared decision-making: a mediation analysis of participation preference," *Health Expectations*, vol. 26, no. 2, pp. 740–751, 2023.

[20] M. Khosravi, Z. Zare, S. M. Mojtabaeian, and R. Izadi, "Artificial intelligence and decision-making in healthcare: a thematic analysis of a systematic review of reviews," *Health services research and managerial epidemiology*, vol. 11, pp. 94–99, 2024.

[21] R. Barbosa, R. Santos, and P. Novais, "Trust-based negotiation in multiagent systems: a systematic review," in *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer, 2023, pp. 133–144.

[22] P. Ji and X. Ma, "A fuzzy intelligent group recommender method in sparse-data environments based on multi-agent negotiation," *Expert Systems with Applications*, vol. 213, p. 119294, 2023.

[23] R. Pereira, T. M. Sousa, T. Pinto, I. Praça, Z. Vale, and H. Morais, "Strategic bidding for electricity markets negotiation using support vector machines," *Trends in Practical Applications of Heterogeneous Multi-agent Systems. The PAAMS Collection*, pp. 9–17, 2014.

[24] F. Fu and H. Zhou, "A combined multi-agent system for distributed multi-project scheduling problems," *Applied Soft Computing*, vol. 107, p. 107402, 2021.

[25] M. Jiang, C. Gao, Y. Yang, and A. K. Pogodaev, "A Novel Containment Control Design Scheme for Second-Order Multi-agent Systems with Adjustable Reference Signals." *IAENG International Journal of Applied Mathematics*, vol. 54, no. 6, pp. 1048–1052, 2024.

[26] R. Aydoğan, D. Festen, K. V. Hindriks, and C. M. Jonker, "Alternating offers protocols for multilateral negotiation," *Modern approaches to agent-based complex automated negotiation*, pp. 153–167, 2017.

[27] T. Baarslag, *Exploring the strategy space of negotiating agents: A framework for bidding, learning and accepting in automated negotiation*. Springer, 2016.

[28] S. Elghamrawy, "Healthcare informatics challenges: A medical diagnosis using multi agent coordination-based model for managing the conflicts in decisions," in *International Conference on Advanced Intelligent Systems and Informatics*. Springer, 2020, pp. 347–357.

[29] T. Sun, Q. Zhu, Y. Xia, and F. Cao, "A bilateral price negotiation strategy based on Bayesian classification and q-learning," *Journal of Information & Computational Science*, vol. 8, no. 13, pp. 2773–2780, 2011.

[30] K. Hindriks and D. Tykhonov, "Opponent modelling in automated multi-issue negotiation using bayesian learning," in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*. Citeseer, 2008, pp. 331–338.

[31] C. Han, Y. Zhang, Q. Chen, T. Jiang, and J. Zhou, "Fault-tolerant Asynchronous Control for FM-II Model-based 2D Markov Jump Systems Under Actuator Failures and Mode Mismatches." *IAENG International Journal of Applied Mathematics*, vol. 54, no. 10, pp. 2083–2091, 2024.

[32] H. Xia, W. Zeng, X. Shen, L. Hu, Y. Jiang, and Y. Wang, "Plant Diseases and Insect Pests Recognition Algorithm Based on D-YOLOv3," in *2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*. IEEE, 2023, pp. 212–216.

[33] M. M. Daud, Z. Kadim, and H. H. Woon, "Detection of Oil Palm Tree and Loose Fruitlets for Fresh Fruit Bunch's Ready-to-Harvest Prediction via Deep Learning Approach," *IAENG International Journal of Computer Science*, vol. 50, no. 4, pp. 1183–1193, 2023.

[34] X. Cai, N. Zhang, G. K. Venayagamoorthy, and D. C. Wunsch II, "Time series prediction with recurrent neural networks trained by a hybrid PSO–EA algorithm," *Neurocomputing*, vol. 70, no. 13-15, pp. 2342–2353, 2007.

[35] T. Baarslag, K. Fujita, E. H. Gerding, K. Hindriks, T. Ito,

N. R. Jennings, C. Jonker, S. Kraus, R. Lin, V. Robu *et al.*, "Evaluating practical negotiating agents: Results and analysis of the 2011 international competition," *Artificial Intelligence*, vol. 198, pp. 73–103, 2013.

[36] P. Faratin, C. Sierra, and N. R. Jennings, "Negotiation decision functions for autonomous agents," *Robotics and Autonomous Systems*, vol. 24, no. 3-4, pp. 159–182, 1998.