



«توسعه‌ی عامل‌های مذاکره‌کننده»

مقدمه

مذاکره، در مدیریت تعامل بین عامل‌های محاسباتی و در هوش مصنوعی توزیع‌شده نقش مهمی ایفا می‌کند. مذاکره‌ی مبتنی بر عامل فرآیند مذاکره را خودکار می‌کند. مذاکره‌ی خودکار فرآیندی است که در آن عامل‌ها به نمایندگی از انسان‌ها در دنیای واقعی برای دستیابی به توافقات سودمند متقابل مذاکره می‌کنند. مذاکره‌ی خودکار شامل عامل‌های مستقلی است که از طرف مذاکره‌کنندگان انسانی در ارتباطات خریدار-فروشنده یا مصرف‌کننده-ارائه‌دهنده به توافقات مشترک دست می‌یابند. تحقیقات در زمینه‌ی مذاکره‌ی خودکار با افزایش توانایی‌های مذاکره‌کنندگان انسانی و کاربردهای بالقوه‌ی آن در حوزه‌های صنعتی و تجاری انجام می‌شود. این فناوری در کاربردهای صنعتی مختلفی از جمله زنجیره تامین، شبکه‌ی هوشمند، تجارت الکترونیک، تخصیص وظایف و رانندگی خودگردان مورد استفاده قرار می‌گیرد.

تعاریف اولیه

یک تابع سودمندی تابعی است که یک مقدار سودمندی $U(\omega_i)$ را به یک خروجی ω_i اختصاص می‌دهد. این توابع، به صورت خصوصی در دسترس هر عامل هستند.

پیشنهاد نهایی با ω^* نمایش داده می‌شود.

هر عامل دارای یک مقدار سودمندی رزرو شده‌ی U_r است، که در صورت عدم توافق آن را دریافت می‌کند. روند یک مذاکره با T نمایش داده می‌شود، که دنباله‌ای از پیشنهادات متناوب توسط عامل‌ها در یک معامله است. و به صورت $T = \{\omega_1^o, \omega_1^s, \dots, \omega_n^o, \omega_n^s\}$ برای یک مذاکره‌ی دوجانبه تعریف می‌شود؛ که ω_i^s و ω_i^o به ترتیب پیشنهادات مربوط به خود عامل و عامل حریف را نشان می‌دهد. همچنین به صورت دو مجموعه‌ی $T^s = \{\omega_1^s, \dots, \omega_n^s\}$ و $T^o = \{\omega_1^o, \dots, \omega_n^o\}$ نیز تعریف می‌شود.

تاریخچه‌ی مذاکرات در دامنه‌ی خاص D با $H_D = \{T_1, \dots, T_k\}$ نشان داده می‌شود.

تنظیمات مذاکره شامل یک پروتکل، دامنه و عامل‌هاست. پروتکل قوانین تعامل را تعیین می‌کند، دامنه فضای نتیجه را تعیین می‌کند و عامل‌ها مذاکره‌کنندگان هستند. مذاکره‌ی دوجانبه نوع رایجی از مذاکره‌ی بین دو مذاکره‌کننده است. پروتکل پیشنهاد متناوب اغلب در مذاکرات دوجانبه استفاده می‌شود. عامل‌ها

سودمندی یک پیشنهاد را بر اساس ارزش هر مسئله و اولویت‌های وزنی آن‌ها ارزیابی می‌کنند. سودمندی با استفاده از یک تابع ارزیابی محاسبه می‌شود. عامل‌ها از استراتژی‌هایی مانند استراتژی‌های وابسته به زمان و رفتار، برای تعیین پذیرش یا عدم پذیرش پیشنهاد استفاده می‌کنند. یک استراتژی شناخته‌شده‌ی وابسته به رفتار، tit-for-tat است.

یادگیری تقویتی مسئله‌ایست که شامل فرآیند تصمیم‌گیری مارکوف (MDP) با حالت‌ها، اقدامات، عملکردهای انتقال و پاداش و یک عامل تخفیف است.

تعدادی از چارچوب‌های پیشنهادی برای توسعه‌ی عامل‌های مذاکره‌کننده

تکنیک‌های یادگیری ماشین را می‌توان به طور موثر در مذاکره‌ی خودکار با استفاده از داده‌های گذشته‌ی خاص-دامنه به کار برد. توابع سودمندی در مذاکره‌های متوالی می‌توانند به دلایل مختلفی تغییر کنند و داشتن یک استراتژی یکسان برای همه چالش برانگیز است. اولین چارچوبی که در ادامه معرفی می‌شود، از یادگیری انتقالی و تشخیص مداوم تغییرات تابع سودمندی برای ایجاد عامل‌های مذاکره‌کننده‌ی تطبیقی استفاده می‌کند؛ که عملکرد آن نسبت به عامل‌های مستقل از دامنه ۶٪ بهتر است. هدف از بررسی این چارچوب، پرداختن به این سوال است که آیا می‌توان یک استراتژی مذاکره‌ی انتها-به-انتها را از داده‌های گذشته آموخت و چگونه می‌توان از آن در هنگام تغییر توابع سودمندی عامل‌ها استفاده کرد. چارچوب پیشنهادی شامل یک معماری یادگیری عمیق ساده شده، روشی برای انطباق با تغییرات در توابع سودمندی با استفاده از یادگیری انتقالی و رویکردی برای شناسایی و اندازه‌گیری چنین تغییراتی است.

۱. رویکرد تطبیقی مبتنی بر یادگیری انتقالی

چارچوب عامل مذاکره‌کننده‌ی خودکار تطبیقی از سه جزء اصلی تشکیل شده است: (۱) مدل پایه، (۲) جریان انطباق مدل و (۳) منتقد؛ همچنین اجزای کمکی مانند پایگاه داده‌های مربوط به ردیابی مذاکره و سیستم موروئی.

(۱) مدل پایه

این مدل یک مدل یادگیری عمیق است که به طور فعال توسط عامل برای استراتژی مناقصه استفاده می‌شود. در ابتدا، با استفاده از داده‌های گذشته آموزش داده می‌شود؛ اما بعداً هر زمان که لازم باشد با مدل‌های تطبیقی جدید جایگزین می‌شود.

برای معماری مدل از یک Long short-term memory (LSTM) دوطرفه، یک لایه‌ی تعبیه‌شده در ابتدا و یک تک لایه‌ی متراکم با فعال‌سازی softmax در انتها استفاده شده است.

برای استفاده از مذاکرات گذشته برای آموزش، از هر $T \in H_D$ یک دنباله‌ی $I_t = \{\omega_1^o, \omega_1^s, \dots, \omega_t^o\}$ از پیشنهادات تا زمان t و یک برچسب $O_t = \{\omega_t^s\}$ به عنوان پیشنهاد خود عامل در زمان t تولید می‌شود.

سپس از تابع هزینه cross-entropy وزن دار با $(U^S(\omega_T^*))^k$ به عنوان وزن هزینه، برای تخصیص وزن بیشتر به روندهای مذاکره با مقدار سودمندی توافق بالاتر، استفاده می‌شود. k یک هایپرپارامتر است.

برخی از پیش پردازش‌ها مانند نگاشت پیشنهادها به یک مقدار عددی و اضافه کردن هر دنباله‌ی ورودی قبل از ارائه‌ی آن‌ها به عنوان ورودی به مدل انجام می‌شود.

اثر بخشی این معماری ساده از این واقعیت ناشی می‌شود که لایه‌ی تعبیه‌شده با موفقیت اطلاعات جزئی را در مورد U_s و U_o ثبت می‌کند.

(۲) جریان انطباق مدل

جریان انطباق مدل در واقع عبارتست از آموزش یک مدل تطبیقی جدید برای استراتژی مناقصه. این جریان شامل دو مرحله است.

در مرحله‌ی اول، از یک سیستم موروثی، به‌طور موقت، به جای مدل پایه برای مذاکرات استفاده می‌شود، که با موفقیت چندین بار با حریف مذاکره می‌کند تا مجموعه‌ای از تعداد کمی از روندهای مذاکره را با توافق‌ها ایجاد کند؛ که آن مجموعه با H_{small} نشان داده می‌شود. علاوه بر این، سیستم موروثی می‌تواند از نمونه‌های منفرد یا چندگانه‌ای از هر نوع عامل مذاکره‌کننده‌ی خودکار سازگار تشکیل شده باشد.

در مرحله‌ی دوم، مدل پایه با استفاده از تکنیک یادگیری انتقالی مبتنی بر اشتراک‌گذاری پارامتر با استفاده از مجموعه داده‌ی تازه‌ی H_{small} برای آموزش، تطبیق داده می‌شود. با تغییر در تابع سودمندی، عملکرد مدل پایه عمده‌تاً به دلیل استفاده از لایه‌ی تعبیه‌شده‌ی قدیمی کاهش می‌یابد. با توجه به این که لایه‌ی تعبیه‌شده اطلاعات جزئی هر دو تابع سودمندی را دریافت می‌کند، ضروری است تا این لایه زمانی که تابع سودمندی تغییر می‌کند، دوباره آموزش داده شود. برای دستیابی به این هدف، پیشنهاد شده که هم لایه‌ی تعبیه‌شده و هم لایه‌ی متراکم مجدداً آموزش داده شوند تا اطلاعات جدید در مورد تابع سودمندی خود عامل و عامل حریف بازیابی شود. با فرض این که استراتژی حریف ثابت است، از وزن‌ها و بایاس‌های لایه‌ی LSTM دوطرفه از مدل پایه مجدداً استفاده خواهد شد. این امر تعداد کل پارامترهای قابل آموزش را کاهش می‌دهد و مدل را قادر می‌سازد تا سریع‌تر و با مقدار داده‌ی کمتری یاد بگیرد. در عمل، به عنوان مرحله‌ی نهایی یادگیری انتقالی، می‌توان کل مدل را با نرخ یادگیری بسیار کم و برای تعداد کمی از دوره‌ها (epochs) تنظیم کرد. اگرچه این منجر به دقت بهتر مدل می‌شود، اما سودمندی مربوط به توافق‌های نهایی در مقایسه با توافق‌های حاصل شده با مدل‌های بدون تنظیم دقیق، تغییراتی قابل توجه از نظر آماری را نشان نمی‌دهند.

پس از ایجاد مدل تطبیقی از طریق این جریان، از مذاکره‌ی بعدی، از این مدل به عنوان مدل پایه استفاده می‌شود و تاریخچه‌ی داده‌های مذاکره‌ی H با H_{small} جایگزین می‌شود. تعداد پارامترهای آموزش نیز نسبت به مدل پایه کم‌تر خواهد بود.

منتقد را می‌توان به‌عنوان یک طبقه‌بند دودویی مدل‌سازی کرد که تعیین می‌کند که آیا عامل باید با مدل پایه ادامه یابد یا جریان انطباق مدل را آغاز کند. به طور واقع‌بینانه‌ای فرض شده که توابع سودمندی خود عامل و عامل حریف برای مجموعه‌ای از مذاکرات متوالی به طور قابل توجهی تغییر نمی‌کند. همچنین برای سادگی، فرض می‌شود که تنها یکی از توابع سودمندی در یک نمونه تغییر می‌کند.

ورودی آن، شامل میانگین توزیع فرکانس پیشنهادات حریف در تاریخچه‌ی مذاکرات و توزیع فرکانس پیشنهادات حریف در آخرین مذاکره (یا تعدادی از مذاکرات آخر حریف) است. و خروجی برابر یکی از دو مقدار ۰ یا ۱ است، که به ترتیب تعیین‌کننده‌ی ادامه دادن عامل با مدل پایه، یا آغاز جریان انطباق مدل هستند.

در این مقاله برای آزمایش این مدل از طبقه‌بند XGBoost استفاده شده است. در هنگام cold start چارچوب یا زمانی که داده‌های کافی برای آموزش یک طبقه‌بند در دسترس نیست، می‌توان از الگوریتم مبتنی بر فاصله‌ی Wasserstein یا انرژی استفاده کرد.

یک گسترش از این منتقد همچنین می‌تواند تغییر در تابع سودمندی خود عامل را نیز در خود جای دهد. یک راه ساده این است که وقتی تابع سودمندی عامل تغییر می‌کند، یک مدل جدید آموزش داده شود. با این حال، تغییرات کوچک در تابع سودمندی ممکن است به هیچ وجه بر عملکرد مدل پایه تأثیری نداشته باشد. برای اندازه‌گیری میزان تغییر میان دو تابع سودمندی، یک معیار جدید بر اساس فاصله‌ی Levenshtein معرفی شد ($\mathcal{L}_{i:j}(U_1, U_2)$):

$$\mathcal{L}_{i:j}(U_1, U_2) = 1 - \frac{L(\Omega_{U_1}[i:j], \Omega_{U_2}[i:j])}{|\Omega_U[i:j]|},$$

$$\text{where } |\Omega_U[i:j]| = |\Omega_{U_1}[i:j]| = |\Omega_{U_2}[i:j]|$$

and L is defined as

$$L(o_1, o_2) = \begin{cases} |o_1| & \text{if } |o_2| = 0, \\ |o_2| & \text{if } |o_1| = 0, \\ L(o_1[1:|o_1|], o_2[1:|o_2|]) & \text{if } o_1[0] = o_2[0], \\ 1 + \min \begin{cases} L(o_1[1:|o_1|], o_2), \\ L(o_1, o_2[1:|o_2|]), \\ L(o_1[1:|o_1|], o_2[1:|o_2|]) \end{cases} & \text{otherwise.} \end{cases}$$

(جمع‌بندی منتقد: یک طبقه‌بندی باینری است که تعیین می‌کند آیا باید با مدل پایه ادامه داد یا جریان انطباق مدل را بر اساس پیشنهادات حریف آغاز کرد. منتقد توزیع فراوانی پیشنهادات حریف در تاریخچه‌ی مذاکره و آخرین روند مذاکره را به عنوان ورودی در نظر می‌گیرد. XGBoost برای آموزش منتقد استفاده می‌شود. همچنین می‌تواند با اندازه‌گیری تغییر بین دو تابع سودمندی با استفاده از فاصله‌ی Levenshtein، تغییرات در تابع خود سودمندی را در خود جای دهد. الگوریتم آن برای تابع سودمندی حریف از یک معیار بر اساس فاصله‌ی Wasserstein یا فاصله‌ی انرژی استفاده می‌کند. و برای تابع سودمندی خود عامل از یک معیار بر اساس فاصله‌ی Levenshtein استفاده می‌کند. منتقد یک مقدار دودویی ۰ یا ۱ را خروجی می‌دهد، که نشان می‌دهد آیا باید با مدل پایه ادامه داد یا جریان

انطباق مدل را آغاز کرد. الگوریتم آن برای تابع سودمندی خود و حریف بررسی می‌کند که آیا مقدار معیار زیر یک آستانه است تا تصمیم بگیرد که آیا با مدل پایه ادامه یابد یا خیر.

ارتباط موثر در مذاکره بسیار مهم است و مهارت‌های ارتباطی زبانی می‌تواند به بیان مقاصد و سرعت بخشیدن به فرآیند کمک کند. بیشتر کارها در زمینه‌ی مذاکره‌ی خودکار، ارتباطات زبانی را نادیده می‌گیرند و بر پیشنهادات متقابل تمرکز می‌کنند. چارچوب بعدی که در ادامه معرفی می‌شود، مهارت‌های ارتباطی زبانی و استراتژی‌های پیشنهادی را با استفاده از شبکه‌های Q عمیق یکپارچه می‌کند، و مذاکره با کانال‌های ارتباطی متعدد را به عنوان یک مشکل تصمیم مارکوف با فضای عمل ترکیبی مدل می‌کند و از یادگیری تقویتی عمیق برای هدایت پیشنهادات بهینه استفاده می‌کند.

محیط مذاکره‌ی در نظر گرفته شده برای به کارگیری این چارچوب، یک سناریوی چندموضوعی دوجانبه است که در آن خریدار و فروشنده با استفاده از پیشنهادات و Cheaptalk با یکدیگر ارتباط برقرار می‌کنند. Cheaptalk به ارتباطات زبانی اشاره دارد که برای منتقل کردن تمایل به مذاکره و سرعت بخشیدن به فرآیند استفاده می‌شود. در حین مذاکره، هر دو طرف به نوبه‌ی خود پیشنهاداتی را ارائه می‌کنند تا زمانی که یکی دیگری را بپذیرد یا به دلیل اتمام زمان، مذاکره خاتمه یابد.

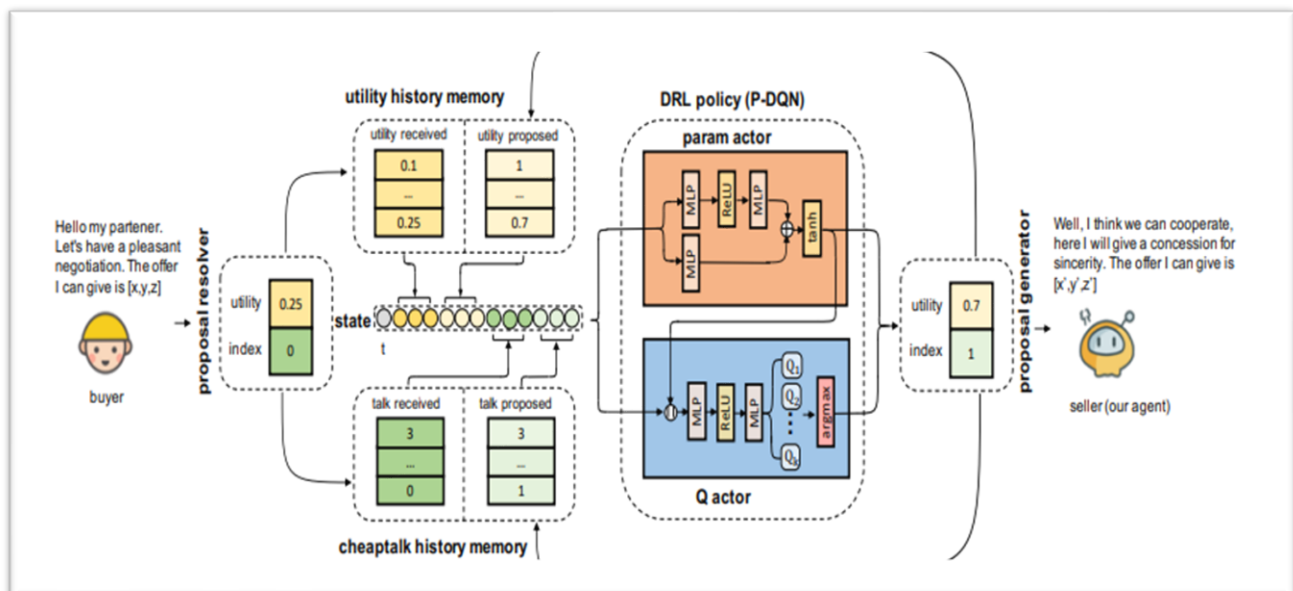
۹ رویداد وجود دارد که در مذاکرات رخ می‌دهند، و در جدول زیر شرح داده شده‌اند. ترکیب این رویدادها و شخصیت‌های مختلف (خوب، سخت)، ۱۸ الگوی Cheaptalk برای مذاکره‌کنندگان تولید می‌کند. عامل می‌تواند در پیشنهاد و Cheaptalk خود، صادق یا ناصدق باشد. برای مثال، دو مورد از رویدادها، عبارتند از: (۱) قبول پیشنهاد حریف و (۲) سلام به حریف برای شروع مذاکره. برای مورد (۱)، بسته به خوب یا بد بودن عامل، به ترتیب، یکی از دو Cheaptalk مثلاً به فرم (۱.۱) "سلام. به نظر من این پیشنهاد واقعا خوب است. من آن را قبول می‌کنم." یا (۲.۱) "اوه. این پیشنهاد به سختی مرا راضی می‌کند؛ اما من آن را می‌پذیرم." رخ خواهد داد.

Negotiation events	Category index
Accept the opponent's proposal	0
Promise to the opponent to make concessions	1
Request opponent to give a concessions	2
Show bottom line to opponent	3
Belief that both sides can get a win-win result	4
Greeting opponent for starting negotiation	5
Punish opponent for breaking its word	6
Forgive opponent for breaking its word	7
Threaten opponent for no conceding	8

۲. رویکرد استفاده از کانال‌های ارتباطی متعدد

چارچوب مذاکره‌ی پیشنهادی به عنوان یک فرآیند تصمیم‌گیری مارکوف (MDP) مدل شده و معماری کلی آن در شکل زیر نشان داده شده‌است. Cheaptalk با شاخص مربوط به خود در قالب‌های از پیش تعریف شده ارائه می‌شود. عامل پیشنهاد را از طرف حریف دریافت می‌کند و از تصمیم‌گیرنده‌ی مربوط به پیشنهاد برای تصمیم‌گیری درباره‌ی پیشنهاد و Cheaptalk استفاده می‌کند. سودمندی

پیشنهاد محاسبه شده و به حافظه‌ی تاریخچه‌ی سودمندی و حافظه‌ی تاریخچه‌ی Cheaptalk اضافه می‌شود. تاریخچه‌ی سودمندی و تاریخچه‌ی Cheaptalk برای ثبت پیشنهادات پیشنهادی و دریافتی در تاریخچه‌ی مذاکره استفاده می‌شود. وضعیت و عمل بر اساس زمان فعلی، تاریخچه‌ی سودمندی و تاریخچه‌ی cheaptalk تعریف می‌شود. هدف عامل به حداکثر رساندن سودمندی پیشنهاد توافق است و تابع پاداش بر این اساس تعریف می‌شود. فرآیند آموزش عامل شامل انتخاب تصادفی یک حریف و به‌روزرسانی وضعیت، انتخاب اقدامات و به‌روزرسانی اطلاعات تاریخچه در طول مذاکره است. پس از چندین دوره آموزش، خط مشی عامل همگرا می‌شود.



تکنیک‌های یادگیری ماشین، مانند یادگیری تقویتی عمیق، در مذاکره‌ی خودکار موفق بوده‌اند، اما در مواجهه با عامل‌های جدید به عنوان حریف، ممکن است با استراتژی‌های ناآشنا یا ناشناخته مواجه شوند که نیازمند یادگیری خط‌مشی‌های جدید از ابتدا باشد. یادگیری انتقالی (TL) به عنوان راه حلی برای تسریع فرآیند یادگیری، استراتژی‌های جدید حریف و در عین حال بهبود عملکرد خط‌مشی‌های آموخته شده پیشنهاد شده است. چارچوب آخری که معرفی می‌شود، مبتنی بر یادگیری انتقالی است که اثربخشی خود را از طریق آزمایش‌های جامع نشان داده است.

۳. رویکرد مبتنی بر یادگیری انتقالی

چارچوب پیشنهادی مبتنی بر یادگیری انتقال برای مذاکره خودکار از سه ماژول تشکیل شده است: مذاکره، تطبیق و انتقال.

(۱) ماژول مذاکره

این ماژول با سایر عامل‌ها تعامل می‌کند، پیشنهادات متقابل ایجاد می‌کند و دربارهی پذیرش یا رد پیشنهادات تصمیم‌گیری می‌کند.

(۲) ماژول تطبیق

این ماژول قابلیت انتقال خطمشی‌ها را با استفاده از دو معیار عملکرد و شباهت اندازه‌گیری می‌کند.

(۳) ماژول انتقال

این ماژول دانش مفید را از خطمشی‌های منبع استخراج می‌کند و برای بهبود عملکرد عامل هنگام مواجهه با عامل‌های جدید به عنوان حریف تصمیم‌گیری می‌کند.

بررسی عملکرد چارچوب‌های معرفی شده و ویژگی‌های مذاکره‌های مناسب برای استفاده از آن‌ها

در ادامه نمونه آزمایش‌های استفاده شده برای بررسی عملکرد چارچوب‌های معرفی شده، شرح داده شده‌اند.

۱. بررسی عملکرد رویکرد تطبیقی مبتنی بر یادگیری انتقالی

۳ دامنه‌ی ANAC 2015 با ۸۸ تابع سودمندی حریف و ۸۸ تابع سودمندی خود عامل برای هر دامنه در نظر گرفته شد. آزمایش‌ها به دو بخش تقسیم شدند: یکی اندازه‌گیری عملکرد کلی چارچوب با تغییر در توابع سودمندی حریف، و دیگری اندازه‌گیری عملکرد کلی چارچوب با تغییر در توابع سودمندی خود عامل. برای آزمایش‌های با تغییر در تابع سودمندی حریف، استراتژی حریف و تابع سودمندی خود عامل ثابت نگه داشته شد. مدل پایه با شبیه‌سازی مذاکرات بین عامل‌ها از سیستم موروثی و استراتژی ثابت حریف آموزش داده شد. سپس ۸۸ مدل تطبیقی با استفاده از جریان انطباق مدل آموزش داده شدند و با سیستم موروثی مقایسه شدند. استراتژی‌های مناقصه‌ی مبتنی بر یادگیری انتقالی بدون استفاده از هیچ استراتژی پذیرشی ارزیابی شدند. توابع سودمندی رزرو شده‌ی U_r برابر ۰ و ضریب تخفیف برابر ۱ در نظر گرفته شد. تابع سودمندی حریف پس از هر ۱۰ دور مذاکره برای واقعی‌تر کردن آزمایش تغییر می‌کرد. برای آزمایش‌های با تغییر در تابع سودمندی خود عامل، تابع سودمندی و استراتژی حریف ثابت نگه داشته شد. ایجاد مدل، تنظیم آزمایش و ارزیابی مشابه آزمایش‌های با تغییر در تابع سودمندی حریف بود. برای آموزش منتقد در شرایطی که تابع سودمندی حریف تغییر می‌کند، ابتدا یک مجموعه داده با دو توزیع پیشنهاد حریف به عنوان ویژگی‌ها و برچسب‌های دودویی ۰ به این معنی که مدل پایه استفاده می‌شود و ۱ به معنای فعال‌سازی جریان انطباق مدل ایجاد شد. پس از آن، از XGBoost برای آموزش یک طبقه‌بند استفاده شد.

نتایج آزمایش با مقایسه‌ی عملکرد مدل‌های تطبیقی با سیستم موروثی در سه حوزه‌ی ANAC 2015 ارائه شده‌اند. مدل‌های تطبیقی مقادیر سودمندی متوسط رو به بالایی را نسبت به سیستم موروثی نشان می‌دهند، که نشان‌دهنده‌ی اثربخشی یادگیری انتقالی مبتنی بر اشتراک‌گذاری پارامتر است. لایه‌ی تعبیه‌شده‌ی مدل‌ها با موفقیت اطلاعات جزئی در مورد توابع سودمندی هر دو عامل را ضبط می‌کند. بازآموزی لایه‌ی تعبیه‌شده برای انطباق با تغییرات در تابع سودمندی ضروری است. عملکرد چارچوب با تنظیمات مختلف منتقد مقایسه می‌شود و مشخص می‌شود که منتقد عملکرد را کاهش نمی‌دهد. منتقد مبتنی بر فاصله‌ی Wasserstein و منتقد مبتنی بر XGBoost عملکرد مشابهی دارند؛ اما پارامتر آستانه نیاز به بهینه‌سازی دارد. چارچوب کلی با یک منتقد هوشمند به میانگین سودمندی متوسط بالاتری نسبت به سیستم موروثی دست می‌یابد. هنگامی که تابع سودمندی خود عامل بیش از حد تغییر می‌کند،

عملکرد مدل پایه آسیب می بیند. مدل تطبیقی بهتر از مدل پایه و مدل موروثی در هر سه حوزه عمل می کند.

۲. بررسی عملکرد رویکرد استفاده از کانال های ارتباطی متعدد

عامل های مذاکره کننده مبتنی بر قانون با ادغام مهارت های ارتباطی زبانی و استراتژی های پیشنهادی طراحی می شوند. رویکرد پیشنهادی با استفاده از الگوریتم P-DQN در برابر عامل های مبتنی بر قانون آموزش داده شد. عامل استفاده کننده از این رویکرد در حدود ۲۰۰۰۰۰ قسمت با پاداش ۰.۶۸ همگرا شد. این عامل در مذاکره با عامل های دوستانه و صادق عملکرد خوبی دارد. در یک مطالعه، از نظر سودمندی متوسط و برنده شدن بهتر از سایر عامل ها است. و مهارت های ارتباطی زبانی و استراتژی های مناقصه را با موفقیت یاد می گیرد. الگوریتم P-DQN به طور قابل توجهی عملکرد عامل را بهبود می بخشد. انواع TP-DQN و OP-DQN الگوریتم P-DQN در مقایسه با ورژن اصلی P-DQN عملکرد ضعیفی دارند. این عامل با دادن امتیاز به حریف ها، صداقت خود را نشان می دهد و از شاخص های خاصی برای ایجاد اعتماد و بیان مودبانه ی نیازها استفاده می کند. نتایج تجربی نشان می دهد که از نظر سودمندی مذاکره و برنده شدن، از عامل های مبتنی بر قانون بهتر عمل می کند.

۳. بررسی عملکرد رویکرد مبتنی بر یادگیری انتقالی

عامل استفاده کننده از این رویکرد در چندین مذاکره با حریفان مختلف و در حوزه های مختلف ارزیابی شد تا توانایی یادگیری کارآمد آن مورد آزمایش قرار گیرد. این عامل از نظر زمان همگرایی و متوسط سودمندی از خط پایه یادگیری از ابتدا بهتر عمل می کند. و از ماژول انتقال برای انتقال دانش از چندین خط مشی منبع استفاده می کند که منجر به همگرایی سریع تر و بهبود عملکرد می شود.

به طور کلی، در مذاکره هایی که توابع سودمندی به سرعت در حال تغییر است، رویکرد اول عملکرد خوبی خواهد داشت. در مذاکره هایی که سریع خاتمه یافتن مذاکره، آگاهی از تمایل حریف به مذاکره اهمیت دارد، و از صادق بودن حریف اطمینان وجود دارد، رویکرد دوم عملکرد خوبی خواهد داشت. در مذاکره هایی که طیف عامل های حریف بسیار گسترده باشد و نیاز به یادگیری های پیاپی و تغییر مداوم خطی مشی هاست، رویکرد سوم عملکرد خوبی خواهد داشت.

بهبود و گسترش رویکردهای معرفی شده

در ادامه نمونه کارهای آینده برای بهبود و توسعه ی چارچوب های معرفی شده، شرح داده شده اند.

۱. گسترش رویکرد تطبیقی مبتنی بر یادگیری انتقالی

می‌توان علاوه بر تطبیق عامل‌ها با تغییرات توابع سودمندی، تغییرات در استراتژی حریف و طراحی اجزای مقاوم در برابر چنین تغییراتی را نیز در نظر گرفت و تطبیق عامل را با توجه به آن‌ها نیز میسر کرد.

۲. گسترش رویکرد استفاده از کانال‌های ارتباطی متعدد

می‌توان مهارت‌های ارتباطی زبانی پیچیده‌تر و همچنین گسترش چارچوب به سایر تنظیمات مذاکره و دیگر مهارت‌های زبانی را در نظر گرفت.

۳. گسترش رویکرد مبتنی بر یادگیری انتقالی

کارهای تحقیقاتی جدیدی از جمله ترکیب تکنیک‌های مدل‌سازی حریف، ارزیابی عملکرد در برابر مذاکره‌کنندگان انسانی و گسترش چارچوب برای مذاکره‌های همزمان، با در نظر گرفتن این رویکرد ممکن می‌شود.

1. A. Sengupta, S. Nakadai, and Y. Mohammad, "Transfer Learning based Adaptive Automated Negotiating Agent Framework," *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, Jul. 2022, doi: 10.24963/ijcai.2022/67.
2. S. Chen and R. Su, "An autonomous agent for negotiation with multiple communication channels using parametrized deep Q-network," *Mathematical Biosciences and Engineering*, vol. 19, no. 8, pp. 7933–7951, Jan. 2022, doi: 10.3934/mbe.2022371.
3. S. Chen, "Transfer Learning based Agent for Automated Negotiation," 2023.