چانه‌زنی هوشمند در بازار با استفاده از یادگیری تقویتی

محمدعلی سعادت‌جو، ولی درهمی و فاطمه سعادت‌جو

چکیده: استفاده از تکنیک‌های فناوری اطلاعات در بازارهای الکترونیکی، پویایی و پیچیدگی سیستم عرضه و تقاضا را بالا برده است. بنابراین به‌کارگیری عامل‌های هوشمند جهت خرید و فروش و چانه‌زنی در این گونه بازارها به‌عنوان یک راهکار مؤثر پیشنهاد شده است. الگوریتم یادگیری تقویتی یکی از روش‌های قوی یادگیری عامل‌هاست که با کمترین اطلاعات ممکن می‌تواند به‌صورت تعاملی برای آموزش عامل، در راستای پیشنهاد قیمت به‌کار گرفته شود. چانه‌زنی یک مذاکره چالش برانگیز و پیچیده است که علت آن تنوع متغیرهای بسیار زیاد در روابط عرضه و تقاضا و دانش ناکافی شرکت‌کنندگان در بازار می‌باشد. در این مقاله نحوه به‌کارگیری یادگیری تقویتی در مسأله چانه‌زنی در دو بازار مناقصه و مزایده در راستای بیشینه‌سازی افزایش سود عامل بیان می‌گردد. متغیرهای حالت، عمل و تابع یادگیری تقویتی برای مسأله چانه‌زنی در بازار به کمک یک مسأله یادگیری تقویتی نمونه فرمول‌بندی می‌شوند. با مقایسه روش ارائه‌شده و یک روش تجربی به این واقعیت خواهیم رسید که عامل آموزش‌دیده، سود به مراتب بیشتری را از یک عامل تجربی کسب می‌نماید.

کلید واژه: **بازار الکترونیکی، چانه‌زنی، یادگیری تقویتی.**

۱- مقدمه

در هزاره سوم، فناوری اطلاعات به‌عنوان عمده‌ترین محور تحول و توسعه در جهان منظور شده است. کاربردهای مختلف فناوری اطلاعات مانند شهر الکترونیکی، دولت الکترونیکی، تجارت الکترونیکی و بازار الکترونیکی با سرعت خیره‌کننده‌ای در حال گسترش می‌باشند. یکی از توسعه‌هایی که در دنیای حاضر بر روی بازارهای الکترونیکی انجام گرفته است، راهبرد انتخاب فروشنده‌ای است که سود حاصل از معامله را برای خریدار در نظام مدیریت زنجیره‌ای عرضه (SCM)[1] بیشینه نماید. نتایج این نظام بر بازدهی مالی و شهرت تجاری تأثیرگذار است. اکثر روندهای انتخاب فروشنده بر اساس مکانیزم مزایده و مذاکره می‌باشد [۱] تا [۳]. به‌طور کلی، تصمیم‌گیری در مذاکره برای انتخاب فروشنده یا خریدار معمولاً به تجربه مدیران وابسته است [۴]. مذاکره‌ها در صنایع اغلب به دلیل تنوع پیش‌زمینه فکری طرفین مذاکره، متغیرهای بسیار درگیر در مزایده/ مناقصه، تعاملات پیچیده و دانش ناکافی شرکت‌کنندگان، کارآمد نمی‌باشند [۵] و [۶]. در یک محیط غیر قطعی، وجود یک نظام هوشمند جهت مذاکره در روند انتخاب فروشنده/ خریدار، با توجه به بالابودن پویایی سیستم و لزوم اتخاذ تصمیم در زمان محدود، برای کاهش وابستگی به تجربه فردی لازم می‌شود. هوشمندبودن روش، منجر به بهبود کیفیت تصمیم‌گیری در مذاکره و کاهش زمان آن خواهد شد. در [۷] یک مدل تصمیم‌گیری مذاکره متوالی پیشنهاد شده است که در بردارنده یک سری از رفتارهای مذاکره هوشمند می‌باشد. یک مدل مذاکره میان واسطه‌های مستقل در [۶] ارائه شده است. این مدل، یک سری از راهبردها و تاکتیک‌هایی را معرفی می‌نماید که واسطه‌ها می‌توانند به‌منظور ارائه و ارزیابی پیشنهادات اولیه از آن استفاده نمایند. در [۸] رویکردی جدید توسط روشی چند- عاملی[2] برای کاهش چشمگیر زمان و هزینه منابع انسانی و به‌منظور ایجاد مذاکره یکپارچه و جامع پیشنهاد شده است. از روش‌های هوشمند در مذاکراتی که هدفشان افزایش توانایی‌های مذاکره‌کننده برای درک طرف مقابل، شناسایی نیازها، محدودیت‌ها، پیش‌بینی عملکرد و افزایش سود ناشی از مذاکره باشد، استفاده می‌شود [۹]. استفاده از تجربیات و تعاملات گذشته، برای افزایش توانایی واسطه‌ها به‌منظور دست‌یابی به نتایج بهتر در [۱۰] مورد توجه قرار گرفته است. در [۱۱] به کمک شبکه‌های عصبی که داده‌های ورودی آن شامل عملکرد، کیفیت، جغرافیا و نرخ انتخاب فروشنده هستند، یک فروشنده با نرخ انتخاب بالاتر از بین سایر فروشندگان انتخاب می‌شود. از طرفی در [۱۱] به واسطه داده‌های آموزش بسیار زیاد استفاده‌شده و پارامترهای معرفی‌شده با توجه به دانش موجود در محیط، به هیچ عنوان مذاکره‌ای برای انتخاب فروشنده صورت نمی‌گیرد و خریدار با توجه به مقادیر موجود که ثابت نیز می‌باشند اقدام به انتخاب فروشنده مورد نظر که بیشترین منفعت را برای او داشته باشد می‌نماید. این در حالی است که در مقاله پیش رو هیچ گونه اطلاعات تکمیلی که کمک به انتخاب فروشنده نماید وجود ندارد و خریدار باید در هر مرحله از مذاکره بر سر قیمت چانه بزند تا این که بتواند فروشنده مورد نظر خود را انتخاب نماید. در [۱۲] به معرفی یک روش مزایده تعاملی فرایند مذاکره انتخاب فروشنده (SSNP)[3] که روشی چالش برانگیز می‌باشد پرداخته شده است و در آن با استفاده از شبکه‌های عصبی، خریدار با توجه به قیمت پیشنهادی مراحل قبلی خود و فروشنده و سایر عوامل محیطی همچون زمان، به پیش‌بینی قیمت هر فروشنده در هر مرحله از چانه‌زنی می‌پردازد و در نهایت با فروشنده‌ای به توافق خواهد رسید که بیشترین سود را برای او به همراه داشته باشد. استفاده از شبکه‌های عصبی برای هوشمندسازی سیستم مذاکره، نیازمند پارامترهای زیادی است تا بتواند اطلاعات کاملی از محیط را در حل مسأله دخالت دهد. این در حالی است که در این مقاله تنها از یک پارامتر قیمت رزروشده استفاده شده است. در اکثر تحقیقات انجام‌شده [۶]، [۷]، [۹]، [۱۰] و [۱۲] به پیش‌بینی قیمت یکی از طرفین مذاکره (خریدار یا فروشنده) پرداخته شده و به پیشنهاد قیمت کمتر توجه شده است. این مسأله منجر به پیش‌بینی نرخ پیشنهادی کنونی در زمان تصمیم‌گیری از طرف یکی از طرفین مذاکره می‌گردد. این موضوع احتمال کنترل بر روند مذاکره شامل مدت زمان هر مذاکره و نرخ توافقی نهایی را افزایش

این مقاله در تاریخ ۱ خرداد ماه ۱۳۹۰ دریافت و در تاریخ ۲۷ خرداد ماه ۱۳۹۱ بازنگری شد.

محمدعلی سعادت‌جو، دانشکده مهندسی برق و کامپیوتر، دانشگاه کاشان، کاشان، (email: saadatjoo@grad.kashanu.ac.ir).
ولی درهمی، دانشکده مهندسی برق و کامپیوتر، دانشگاه یزد، یزد، (email: vderhami@yazduni.ac.ir).
فاطمه سعادت‌جو، گروه مهندسی کامپیوتر، مؤسسه آموزش عالی جهاد دانشگاهی استان یزد، یزد، (email: fatemehsaadatjoo@gmail.com).

1. Supply Chain Management
2. Multi - Agents
3. Supplier Selection Negotiation Process

Archive of SID
www.SID.ir

می‌دهد و همین امر هوشمندبودن مذاکره را تحت‌الشعاع قرار می‌دهد. از طرفی در مسایل کاربردی، لازم است تا پیش‌بینی قیمت طرف مقابل به‌عنوان اقدام اولیه برای تعیین قیمت باشد و بدین ترتیب مذاکره در شرایطی واقعی‌تر ادامه پیدا کند.

در این مقاله، عامل به‌صورت مستقیم یاد می‌گیرد که چه قیمتی را پیشنهاد بدهد. مدل پیشنهاد قیمت از یادگیری تقویتی بهره می‌برد و متناسب با نوع بازار انتخابی مزایده/مناقصه به انتخاب خریدار/فروشنده می‌پردازد. روش یادگیری تقویتی نسبت به سایر روش‌های یادگیری سربار کمتری دارد و به کمک آن می‌توان به آموزش عامل هوشمند برای چانه‌زنی به‌صورت خودکار پرداخت. با استفاده از این روش زمان لازم برای انتخاب خریدار/ فروشنده کاهش یافته و سود حاصل از معامله برای طرف هوشمند معامله افزایش می‌یابد. الگوریتم یادگیری تقویتی بر مبنای تعامل با محیط می‌باشد که تنها با استفاده از یک معیار اسکالر راندمان که سیگنال تقویت یا جایزه نامیده می‌شود، بدون نیاز به سرپرست قادر به آموزش عامل‌ها در محیط‌های پیچیده، غیر قطعی و تصادفی به‌صورت برخط می‌باشد [۱۳]. چالش‌های به‌کارگیری یادگیری تقویتی در هر مسئله، شامل انتخاب عمل مناسب، تعریف حالت‌ها و تعریف تابع یادگیری تقویتی می‌باشد که در این تحقیق راهکار مناسب برای موارد مذکور در مسأله بازار ارائه شده است.

ساختار این مقاله به شرح زیر می‌باشد: در بخش دوم الگوریتم یادگیری تقویتی معرفی می‌گردد. چارچوب روند انتخاب خریدار/ فروشنده در مزایده/ مناقصه در بخش سوم می‌آید. در بخش چهارم الگوریتم پیشنهادی در بازار مزایده/ مناقصه بیان شده است. نتایج شبیه‌سازی‌ها در بخش پنجم و نهایتاً در بخش آخر نتیجه‌گیری آمده است.

۲- یادگیری تقویتی

به معنای آموزش آنچه باید انجام شود- چگونگی نگاشت وضعیت‌ها به عمل- برای آن که یک سیگنال عددی (جایزه) ماکزیمم شود. در یادگیری تقویتی، یادگیرنده و تصمیم‌گیرنده را عامل و چیزی که عامل با آن تعامل دارد شامل هر چیز خارج از عامل، محیط نامیده می‌شود. در هر قدم زمانی، عامل حالت جاری s_t را مشاهده نموده و عمل a_t را از مجموعه عمل‌های ممکن با توجه به سیاست اتخاذشده انتخاب کرده و به محیط اعمال می‌نماید. محیط با احتمال $p(s_t, a_t, s_{t+1})$ به حالت s_{t+1} می‌رود، همچنین محیط سیگنال تقویت $r(s_t, a_t)$ را که به‌صورت r_{t+1} هم نشان داده می‌شود توسط عامل، دریافت می‌نماید [۱۴].

قانونی که عامل با توجه به آن در هر حالت، عمل را برای اجرا انتخاب می‌کند، سیاست می‌نامند و معمولاً به‌صورت $\pi(s,a)$ نشان داده می‌شود که در آن s بیانگر حالت و a بیانگر عمل است. هدف عامل، انجام عمل‌هایی است که به حالت‌هایی با بالاترین ارزش برسد نه به بالاترین پاداش، چرا که جایزه‌ها مطلوبیت لحظه‌ای و ارزش مطلوبیت در بلندمدت را نشان می‌دهد. تابع ارزش برابر امید حاصل جمع کاهش‌یافته سیگنال‌های تقویت دریافت‌شده در طول زمان می‌باشد. به عبارت دیگر ارزش یک حالت، کل مقدار پاداشی است که عامل می‌تواند بعد از شروع از آن نقطه، انتظار دریافت داشته باشد. تابع ارزش حالت تحت سیاست π به‌صورت $V^{\pi}(s)$ نشان داده می‌شود

$$V^{\pi}(s) = E_{\pi}\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s\} \quad , \quad 0 \le \gamma \le 1 \qquad (1)$$

که در آن γ فاکتور تخفیف و $E_{\pi}\{\}$ امید ریاضی است. به‌طور مشابه ارزش عمل a در حالت s تحت سیاست π را تابع ارزش- عمل نامیده و با $Q^{\pi}(s,a)$ نشان می‌دهند

$$Q^{\pi}(s,a) = E_{\pi}\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s \ , \ a_t = a\} \qquad (2)$$

یکی از مهم‌ترین پیشرفت‌ها در یادگیری تقویتی توسعه، الگوریتمی مستقل از سیاست بود که یادگیری Q نامیده شد. این روش به تخمین مقادیر تابع ارزش- عمل می‌پردازد. مقادیر ارزش- عمل در روش مذکور به‌صورت زیر تازه‌سازی می‌گردند [۱۵]

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha_t \\ [r_{t+1} + \gamma \max_{b \in A} Q(s_{t+1}, b) - Q(s_t, a_t)] \qquad (3)$$

که α نرخ یادگیری، s_t حالت، γ فاکتور تخفیف و r مقدار جایزه آنی است. عبارت سوم در فرمول فوق $(\max_{b \in A} Q(s_{t+1}, b))$ برابر با مقدار ارزش حالت بعدی $V(s_{t+1})$ می‌باشد. در [۱۵] سه روش غیر مستقیم برای انتخاب عمل معرفی شده است. از میان این سه روش، در این مقاله روش بیشینه نرم با توجه به این مزیت که ارتباط بین احتمال انتخاب عمل‌ها با مقادیر ارزش- عمل در راستای انتخاب عمل بهینه می‌باشد، به کار گرفته شده است. این روش یکی از کاربردی‌ترین روش‌های انتخاب عمل می‌باشد. فرمول احتمال انتخاب هر عمل در این روش به‌صورت زیر تعریف می‌گردد

$$P(a) = \frac{\exp(Q(s,a)/T)}{\sum_{b \in A} \exp(Q(s,b)/T)} \qquad (4)$$

که $T > 0$ ضریب دما نامیده می‌شود. همچنان که از فرمول مشخص است اگر $T \to \infty$، انتخاب‌ها کاملاً تصادفی می‌گردند. معمولاً در ابتدای یادگیری مقدار ضریب دما بزرگ و در حین یادگیری هر چه به سمت جلو می‌رویم مقدار ضریب دما کاهش می‌یابد تا از تجربیات قبلی بیشتر استفاده گردد.

۳- فرمول‌بندی مسئله

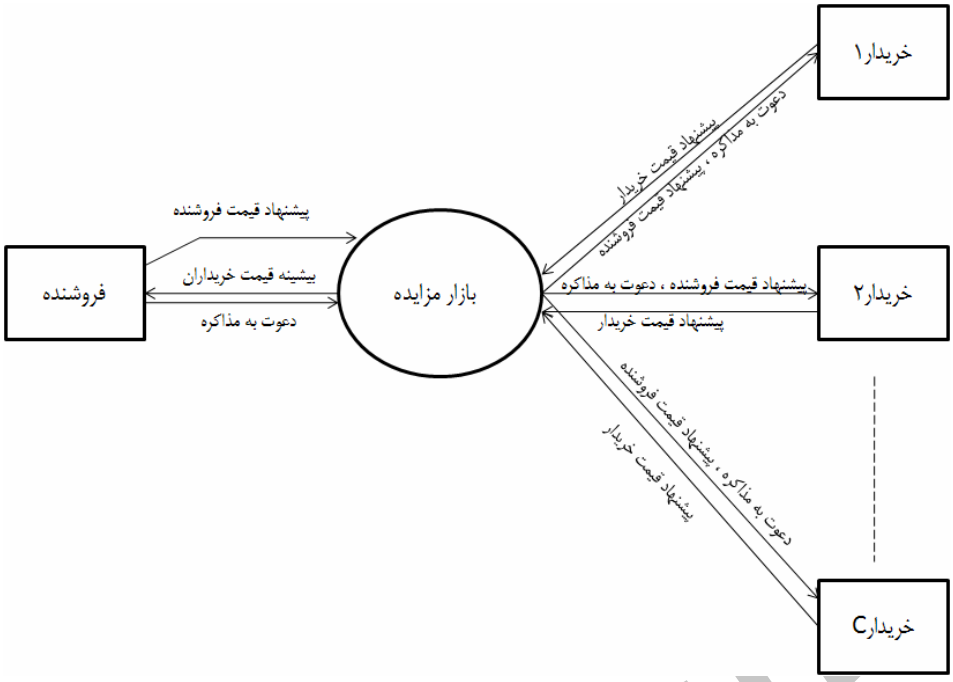
در این بخش روند عملکرد بازار در مذاکره تعاملی [۱۲] به‌صورت دو مدل مجزا شرح داده می‌شود. یکی از این مدل‌ها بازار مزایده را برای انتخاب خریدار و دیگری بازار مناقصه را برای انتخاب فروشنده مورد ارزیابی قرار می‌دهد. این دو مدل از نظر شرایط آغاز و پایان معامله، حداکثر زمان معامله و اجزای چارچوب معامله با یکدیگر یکسان هستند. در ادامه به بیان این شباهت‌ها پرداخته می‌شود.

اولین مرحله از مذاکره در مزایده/ مناقصه با پیشنهاد قیمت توسط فروشنده آغاز می‌شود. هر یک از طرفین معامله، با توجه به قیمت پیشنهادی طرف مقابل، قیمتی را پیشنهاد می‌دهد. عامل زمان در روند مذاکره تعاملی از مهم‌ترین و تأثیرگذارترین عوامل می‌باشد. معامله باید در یک زمان محدود پایان یابد و این زمان با $t^{\max}$ نشان داده می‌شود. مذاکره زمانی پایان می‌پذیرد که یکی از سه شرط زیر برآورده شده باشد:

۱) یکی از طرفین مذاکره (خریدار/فروشنده)، قیمتی را پیشنهاد نماید که قیمت رزرونموده خود را نادیده بگیرد. برای توضیح بیشتر μ_c را به‌عنوان قیمت رزروشده برای خریدار در نظر می‌گیریم. این قیمت در حقیقت، حداکثر مبلغی است که هر خریدار می‌تواند در بازار مزایده پیشنهاد دهد. به‌صورت متناظر قیمت رزروشده برای فروشنده با μ_s نشان داده می‌شود. این قیمت حداقل قیمتی است که فروشنده می‌تواند کالای خود را به فروش رساند. در صورتی که خریدار از قیمت رزروشده خود عبور نماید و قیمتی بالاتر از آن را پیشنهاد نماید و یا فروشنده قیمتی پایین‌تر از قیمت رزروشده خود پیشنهاد دهد، مذاکره به شکست می‌انجامد ($p_c(t) > \mu_c$ و یا

Archive of SID

www.SID.ir



شکل ۱: چارچوب روند انتخاب خریدار در مزایده.

$p_s(t) < \mu_s$). $p_c(t)$ و $p_s(t)$ به‌ترتیب قیمت پیشنهادی خریدار و فروشنده در مرحله tام مذاکره می‌باشد. با برقراری این شرط، توافقی بین طرفین مذاکره حاصل نمی‌گردد. این حالت زمانی اتفاق می‌افتد که یکی از طرفین مذاکره (خریدار یا فروشنده) پیشنهاداتی حریصانه در راستای افزایش سود خود ارائه نمایند.

۲) اگر مدت زمان مذاکره از $t^{\max}$ بیشتر شود، مذاکره بدون دست‌یابی به هیچ نتیجه‌ای پایان می‌یابد.

۳) اگر قیمت پیشنهادی فروشنده در مرحله tام از قیمت پیشنهادی خریدار در مرحله $t-۱$ام کمتر شود ($p_s(t) < p_c(t-۱)$)، در این صورت مذاکره با انتخاب قیمت فروشنده در مرحله $t-۱$ام به‌عنوان قیمت توافقی پایان می‌پذیرد.

چارچوب روند مذاکره انتخاب خریدار شامل سه جزء مجموعه‌ای از خریداران، بازار مزایده و یک فروشنده می‌باشد. چارچوب مذاکره انتخاب فروشنده شامل مجموعه‌ای از فروشندگان، بازار مناقصه انتخاب فروشنده و یک خریدار است. تنها تفاوت بین چارچوب‌های فوق در تعداد خریداران و فروشندگان در دو مدل مزایده و مناقصه می‌باشد که در ادامه به بررسی جزئیات بیشتر آنها خواهیم پرداخت.

۳-۱ چارچوب روند انتخاب خریدار در مزایده

چارچوب روند مذاکره برای انتخاب خریدار در شکل ۱ نمایش داده شده است. این چارچوب شامل سه جزء مجموعه‌ای از خریداران، بازار مزایده و یک فروشنده می‌باشد. خریداران، سفارشات خود را در دوره زمانی تعیین‌شده به فروشنده ارائه می‌دهند. این چارچوب با استفاده از ظرفیت موجود انبار، سفارشات برنامه‌ریزی شده‌ای را جهت مذاکره برای انتخاب یک خریدار مناسب در هر سفارش ایجاد می‌کند.

در [۱۲] قیمت پیشنهادی برای خریداران در مرحله tام به‌صورت زیر تولید می‌گردد

$$P_c(t) = [k_c(t) + ۱ - k_c(t)(\frac{t-۱}{t^{\max}-۱})^{\frac{۱}{\beta_c}}]\mu_c, \quad t = ۱,۲,\ldots,t^{\max} \qquad (۵)$$

$k_c(t) \geq ۰$ و β_c دو پارامتر تصمیم‌گیری خریداران هستند و همواره $k_c(t)$ است. μ_c نشان‌دهنده قیمت رزروشده برای خریدار و β_c کنترل‌کننده منحنی نرخ پیشنهادی خریدار با توجه به قیمت پیشنهادشده توسط فروشنده در طول مذاکره می‌باشد. در واقع این پارامتر، سرعت میزان افزایش یا کاهش قیمت خریدار را کنترل می‌کند. زمان جاری در مذاکره با t و حداکثر تعداد مجاز دفعات مذاکره با $t^{\max}$ نشان داده می‌شود. به‌منظور هوشمندسازی عامل خریدار در راستای ارائه قیمتی واقعی‌تر، در (۵) به‌جای $۱/\beta_c$، اختلاف قیمت پیشنهادی فروشنده و خریدار، نسبت به قیمت خریدار در مرحله $t-۱$ام از روند مذاکره در نظر گرفته شده است. لذا فرمول پیشنهاد قیمت هر خریدار در مرحله tام به‌صورت زیر می‌باشد

$$P_c(t) = [k_c(t) + ۱ - k_c(t)(\frac{t-۱}{t^{\max}-۱})^{۱-\frac{p_s(t-۱)}{p_c(t-۱)}}]\mu_c, \quad t = ۱,۲,\ldots,t^{\max} \qquad (۶)$$

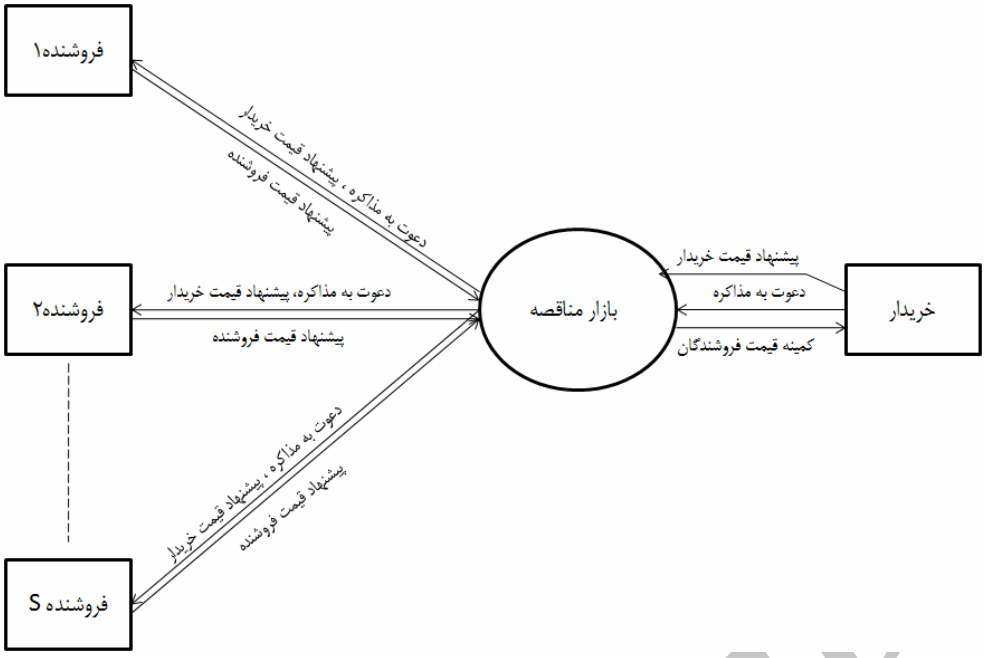
بعد از این که تمام خریداران در مرحله tام قیمت پیشنهادی خود را ارائه نمودند، خریداری انتخاب می‌شود که بیشترین قیمت را پیشنهاد داده باشد. همین امر باعث می‌شود تا در صورت انجام معامله، خریداری به‌عنوان برنده مزایده انتخاب شود که بیشترین سود را برای فروشنده به همراه داشته است. استراتژی پیشنهاد قیمت توسط فروشنده، مسأله‌ای است که در این مقاله به آن پرداخته می‌شود و در بخش چهارم روشی مبتنی بر یادگیری تقویتی برای آن ارائه می‌گردد.

۳-۲ چارچوب روند انتخاب فروشنده در مناقصه

در این بخش روند و چارچوب مذاکره انتخاب فروشنده مورد بررسی قرار می‌گیرد. این چارچوب در شکل ۲ نمایش داده شده است. اجزای این چارچوب شامل مجموعه‌ای از فروشندگان، بازار مناقصه انتخاب فروشنده و یک خریدار می‌باشد. قیمت پیشنهادی فروشنده با توجه به ظرفیت موجود در انبار، برای هر سفارش داده می‌شود و خریدار با توجه به کمینه قیمت فروشندگان، قیمتی را پیشنهاد می‌دهد. موضوع مذاکره تنها قیمت است و در پایان مذاکره، فروشنده‌ای انتخاب می‌شود که بتواند سفارش و زمان مقتضی مطلوب را برای خریدار ایجاد نماید.

Archive of SID

www.SID.ir



شکل ۲: چارچوب روند انتخاب فروشنده در مناقصه.

در [۱۲] نحوه محاسبه قیمت پیشنهادی فروشندگان در مرحله tام به‌صورت زیر معرفی شده است

$$P_s(t)=[1+k_s(t)-k_s(t)(\frac{t-1}{t^{\max}-1})^{\frac{1}{\beta_s}}]\mu_s \quad (7)$$
$$, \quad t=1,2,\ldots,t^{\max}$$

$k_s(t)$ و β_s، دو پارامتر تصمیم‌گیری فروشندگان می‌باشند که همواره $k_s(t)\geq 0$ است. μ_s نشان‌دهنده قیمت رزروشده برای فروشنده و β_s کنترل‌کننده منحنی نرخ پیشنهادی فروشنده با توجه به قیمت پیشنهادشده توسط خریدار در طول مناقصه می‌باشد. به‌منظور هوشمندسازی عامل فروشنده در راستای ارائه قیمتی واقعی‌تر، در (۷) به‌جای $1/\beta_s$، اختلاف قیمت پیشنهادی خریدار و فروشنده، نسبت به قیمت فروشنده در مرحله $t-1$ام از روند مذاکره در نظر گرفته شده است. لذا فرمول پیشنهاد قیمت فروشنده استفاده‌شده در بازار مناقصه به‌صورت زیر در نظر گرفته شده است

$$P_s(t)=[1+k_s(t)-k_s(t)(\frac{t-1}{t^{\max}-1})^{\frac{p_s(t-1)-p_c(t-1)}{P_s(t-1)}}]\mu_s \quad (8)$$
$$, \quad t=1,2,\ldots,t^{\max}$$

از بین تمام فروشندگانی که در مرحله tام قیمتی را پیشنهاد می‌دهند، فروشنده‌ای برگزیده می‌شود که کمترین قیمت را ارائه داده باشد. با اتخاذ این روال، فروشنده‌ای به‌عنوان برنده مناقصه انتخاب می‌شود که بیشترین سود را برای خریدار به همراه داشته باشد. استراتژی پیشنهاد قیمت توسط خریدار با استفاده از یادگیری تقویتی در بخش بعد توضیح داده خواهد شد.

۴- استفاده از یادگیری تقویتی در چانه‌زنی هوشمند

در این بخش، با استفاده از یادگیری تقویتی به یادگیری و هوشمندسازی عامل‌ها در بازار مزایده/ مناقصه پرداخته می‌شود.

۴-۱ مدل پیشنهاد قیمت فروشنده/ خریدار بر اساس یادگیری تقویتی

همان طور که قبلاً بیان شد، یکی از چالش برانگیزترین موارد استفاده از یادگیری تقویتی نحوه تعریف فضای حالت و عمل است. در ادامه ابتدا به نحوه انتخاب حالت‌ها و عمل می‌پردازیم. حالت‌ها به‌صورت دوبعدی $x(t)=(x_1(t),x_2(t))$ تعریف می‌گردند. بعد اول بردار حالت $x_1(t)$، اختلاف بین قیمت انتخاب‌شده پیشنهادی فروشنده و خریدار در معامله می‌باشد که با توجه به نوع معامله به‌صورت زیر نمایش داده می‌شود

$$x_1(t)=\begin{cases} p_s(t)-\max_{i=1}^{M} p_{ic}(t) \\ \min_{i=1}^{N} p_{is}(t)-p_c(t) \end{cases} \quad (9)$$

که $\max_{i=1}^{M} p_{ic}(t)$ به‌عنوان بیشینه قیمت پیشنهادی خریداران در بازار مزایده و M تعداد خریداران در نظر گرفته شده و $\min_{i=1}^{N} p_{is}(t)$ کمینه قیمت پیشنهادی فروشندگان در بازار مناقصه و N تعداد فروشندگان می‌باشد. بعد دوم بردار حالت برای بازارهای مزایده/ مناقصه، اختلاف بین قیمت‌های پیشنهادی فروشنده و خریدار در دو مرحله متوالی معامله است و به‌صورت زیر در نظر گرفته می‌شود

$$x_2(t)=x_1(t)-x_1(t-1) \quad (10)$$

عمل انتخابی، مقدار قیمتی است که در هر مرحله از معامله توسط فروشنده/ خریدار پیشنهاد می‌شود. از رابطه زیر برای محاسبه میزان کاهش/ افزایش قیمت فروشنده/ خریدار در هر مرحله از زمان مذاکره استفاده می‌گردد

$$a(t)=a_price(t)\times\Delta \quad (11)$$

که در آن $a(t)$ عمل انتخاب‌شده فروشنده با استفاده از روش بیشینه نرم در مرحله tام مزایده و $a_price(t)\in\{1,2,\ldots,n\}$ می‌باشد که n تعداد عمل‌های انتخابی ممکن برای عامل است. با توجه به این که در بازار مزایده، هدف هوشمندسازی فروشنده می‌باشد و فروشنده باید در هر مرحله از چانه‌زنی قیمت خود را کاهش دهد تا با خریدار به توافق برسد، لذا مقدار ثابت و منفی Δ در عدد عمل انتخابی ضرب می‌گردد تا میزان کاهش قیمت فروشنده را در لحظه tام مزایده مشخص نماید. از طرفی در بازار مناقصه، هدف هوشمندسازی عامل خرید می‌باشد و در هر مرحله از چانه‌زنی لازم است تا قیمت پیشنهادی خریدار افزایش یابد. بنابراین Δ مقداری مثبت در نظر گرفته می‌شود و با انتخاب هر عمل، مبلغی به قیمت پیشین خریدار با توجه به میزان قیمت فروشنده اضافه می‌گردد. با

Archive of SID
www.SID.ir

جدول ۱: مقایسه الگوریتم پیشنهادی و روش تجربی در مزایده/ مناقصه.

نوع معامله	نام روش	متوسط قیمت توافقی	تعداد شکست
مناقصه	روش تجربی	۱۵٫۶۲	۶
	روش پیشنهادی ما (RL)	۱۱٫۹۹	۲
مزایده	روش تجربی	۸۳٫۱۴	۴
	روش پیشنهادی ما (RL)	۹۲٫۵۴	۲

توجه به توضیحات داده‌شده می‌توان فرمول زیر را برای محاسبه قیمت هر فروشنده/ خریدار به‌دست آورد

$$\begin{cases} p_s(t) = p_s(t-1) + a(t) \\ p_c(t) = p_c(t-1) + a(t) \end{cases} \quad (12)$$

که در آن $p_c(t)$ قیمت پیشنهادی خریدار در مناقصه و $p_s(t)$ قیمت پیشنهادی فروشنده در مرحله tام مزایده است. این قیمت‌ها با توجه به عمل انتخابی و قیمت پیشنهادی در مرحله قبل در مذاکره می‌باشند. در شروع مذاکره $(t=1)$، قیمت پیشنهادی فروشنده/خریدار از فرمول زیر محاسبه می‌گردد

$$\begin{cases} p_s(1) = \max_{i=1}^{M} p_{ic}(1) \times 3 + a(1) \\ p_c(1) = \min_{i=1}^{N} p_{is}(1) / 3 + a(1) \end{cases} \quad (13)$$

مهم‌ترین نکته در یادگیری تقویتی، تعریف مناسب تابع سیگنال تقویت (مقدار جایزه آنی $r(t)$) می‌باشد. در هر مرحله از مزایده/مناقصه مقدار جایزه آنی به‌صورت زیر در نظر گرفته می‌شود

$$\begin{cases} 1 \\ r(t) = \dfrac{(\max(x_1) - \min_{i=1}^{n} p_s(t) - p_c(t))}{\max(x_1)} \\ -1 \end{cases} \quad (14)$$

که در آن بخش اول در لحظه‌ای که معامله انجام می‌شود، بخش دوم در حین مذاکره و بخش سوم در زمان شکست است.

۴-۲ الگوریتم پیشنهاد قیمت در مزایده/ مناقصه

روند پیشنهاد قمیت در بازار به‌صورت زیر می‌باشد:

در لحظه شروع $t=1$، به‌ازای هر فروشنده/خریدار، μ_s با توجه به سطح انبار هر فروشنده و μ_c با توجه به درآمد هر خریدار به‌صورت ثابت تعریف می‌شوند. با توجه به نوع مذاکره، $p_s(1)$ و $p_c(1)$ از (۱۳) به‌دست می‌آیند. لازم به ذکر است که اولین عمل به‌صورت تصادفی انتخاب می‌شود.

برای مراحل بعد، $t=t+1$ در نظر گرفته می‌شود و اگر $t>t^{\max}$ باشد، مذاکره به شکست می‌انجامد.

در مزایده، قیمت پیشنهادی هر خریدار بر اساس (۶) و قیمت پیشنهادی فروشنده از بخش اول (۱۲) به‌دست می‌آید. در مناقصه، قیمت پیشنهادی هر فروشنده از (۸) و قیمت پیشنهادی خریدار از بخش دوم (۱۲) محاسبه می‌شود. سپس از (۱۴) سیگنال تقویتی محاسبه شده و به کمک آن مقادیر ارزش- عمل هر یک از حالت‌ها با توجه به عمل انتخابی به‌هنگام می‌شود.

در مزایده اگر قیمت پیشنهادی خریدار از قیمت رزروشده او بیشتر باشد $(p_c(t) > \mu_c)$، و در مناقصه اگر قیمت پیشنهادی فروشنده کمتر از سطح انبار باشد $(p_s(t) < \mu_s)$، مذاکره به شکست می‌انجامد.

در صورتی که قیمت پیشنهادی فروشنده در مرحله tام کمتر از قیمت پیشنهادی خریدار در مرحله $t-1$ام باشد $((p_s(t) < p_c(t-1))$، توافق حاصل شده و مذاکره متوقف می‌شود و در غیر این صورت، اگر شرایط انجام مذاکره برقرار باشد، چانه‌زنی ادامه می‌یابد.

۵- شبیه‌سازی

جهت ارزیابی روش ارائه‌شده، مقایسه‌ای بین الگوریتم پیشنهادی با روش تجربی در شرایط محیطی یکسان به کمک نرم‌افزار MATLAB صورت می‌گیرد. در شبیه‌سازی فعلی حداکثر تعداد مجاز دفعات مذاکره $t^{\max}=15$ در نظر گرفته شده است. برای پیشنهاد قیمتی با بالاترین سود ممکن عملیات چانه‌زنی برای ۱۰۰۰ مذاکره انجام شده است. مقادیر ارزش- عمل در شروع اولین اجرا با صفر مقداردهی شده‌اند. در شبیه‌سازی انجام‌شده، پارامترهای ضریب یادگیری اولیه برابر $\alpha=0.6$، فاکتور تخفیف $\gamma=0.95$ و مقدار اولیه ضریب دما در (۴) برابر $T_0=70$ تعیین می‌گردند. هر دو روش برای خرید/فروش ۵۰ محصول به‌صورت مستقل اجرا شده‌اند. نتایج ارائه‌شده با میانگین‌گیری از نتایج حاصل از هر اجرا می‌باشد. در مزایده، قیمت رزروشده برای هر فروشنده متناسب با سطح انبار تعیین می‌شود. در اینجا $\mu_s=10$ و تعداد خریداران $M=10$ در نظر گرفته شده است. برای مناقصه تعداد فروشندگان $N=10$ و قیمت رزروشده برای خریدار $\mu_c=30$ انتخاب شده است که قیمت رزروشده برای هر خریدار متناسب با توانایی مالی او تعیین می‌گردد.

در مزایده به روش تجربی با اقتباس از فرمول فروش آمده در [۱۲]، فروشنده، قیمت خود را با توجه به فرمول زیر پیشنهاد می‌دهد

$$p_s(t) = \frac{p_s(t-1) + \max_{i=1}^{M} p_{ic}(t)}{3} - \sigma \quad (15)$$

و قیمت پیشنهادی خریدار با توجه به (۶) به‌دست می‌آید.

در مناقصه به روش تجربی، فروشنده از (۸) برای پیشنهاد قیمت بعدی خود استفاده می‌کند و در مناقصه به روش تجربی با اقتباس از فرمول خرید آمده در [۱۲]، خریدار با توجه به فرمول زیر قیمت پیشنهادی خود را بیان می‌دارد

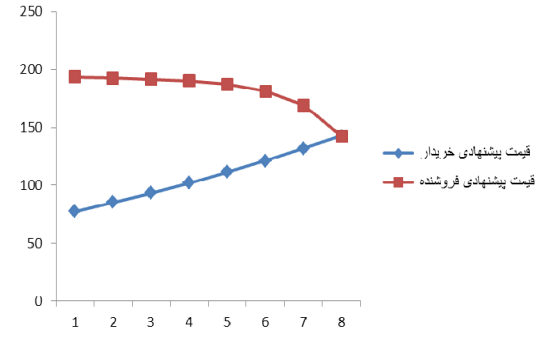
$$p_c(t) = \frac{p_c(t-1) + \min_{i=1}^{N} p_{is}(t)}{7} + \sigma \quad (16)$$

که در آن σ عددی ثابت است و توسط طراح تعیین می‌گردد. برای مقایسه دو روش تجربی و پیشنهادی، در شروع هر مذاکره شرایط یکسانی در نظر گرفته می‌شود. برای انجام این کار در لحظه $t=1$، قیمت پیشنهادی $p_c(1)$ برای کلیه خریداران و قیمت پیشنهادی $p_s(1)$ برای فروشنده در بازار مزایده در هر دو روش تجربی و پیشنهادی یکسان و برابر می‌باشند و این امر برای بازار مناقصه نیز برقرار است. نتایج شبیه‌سازی در جدول ۱ نشان داده شده است. در این جدول، ستون اول نشان‌دهنده نوع معامله، ستون دوم معرف نام روش، سومین ستون بیانگر میانگین قیمت توافقی و ستون آخر تعداد شکست‌ها در ۵۰ اجرا می‌باشد. همان طور که از این جدول پیداست، میانگین قیمت توافقی در روش پیشنهادی بر اساس یادگیری تقویتی از روش تجربی، در بازار مزایده بیشتر و در بازار مناقصه کمتر است که بیانگر کسب سود بیشتر در روش پیشنهادی است. همچنین تعداد شکست‌ها در روش تجربی به مراتب بیشتر از روش پیشنهادی است.

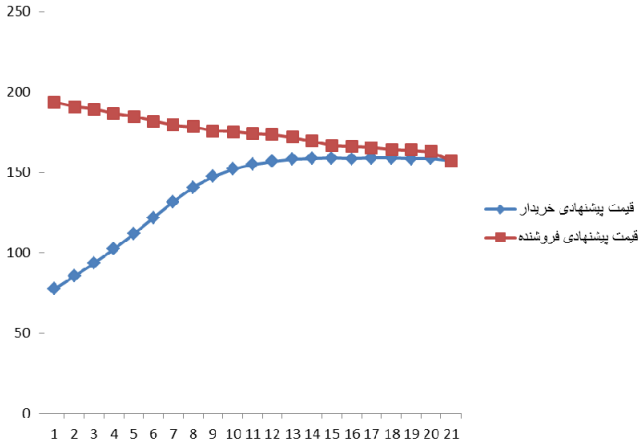
در شکل‌های ۳ و ۴ قیمت پیشنهادی هر دو روش در طول مدت انجام یک مزایده در یکی از اجراها نشان داده شده‌اند. همان طور که از شکل ۳

Archive of SID

www.SID.ir



شکل ۳: روند پیشنهاد قیمت در روش تجربی.



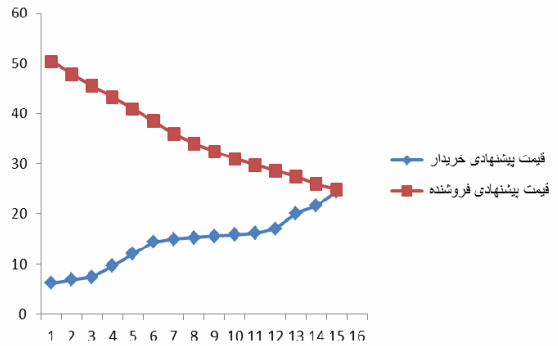
شکل ۴: روند پیشنهاد قیمت در روش پیشنهادی.

پیداست، قیمت توافقی در روش تجربی برابر ۱۴۲٫۶۳ می‌باشد. فروشنده در روش تجربی بسیار کورکورانه و حریصانه به پیشنهاد قیمت می‌پردازد، بدون این که در هر پیشنهاد، قیمت خریدار را در قیمت خود دخیل نماید. لذا فروشنده بدون در نظر داشتن افزایش سود خود به معامله می‌پردازد. در شکل ۴ معامله‌ای با قیمت پیشنهادی آغازین برابر برای خریداران و فروشنده و با استفاده از یادگیری تقویتی نشان داده شده است. از این شکل در می‌یابیم که قیمت توافقی در روش پیشنهادی ۱۵۸٫۰۳ است که حدود ۱۱٪ نسبت به روش تجربی افزایش یافته است. به بیان دیگر سود حاصل از معامله برای خریدار به همین درصد افزایش یافته است. اگر چه عامل هوشمند برای رسیدن به توافق، زمان بیشتری را صرف نموده ولی سود بیشتری را کسب کرده است.

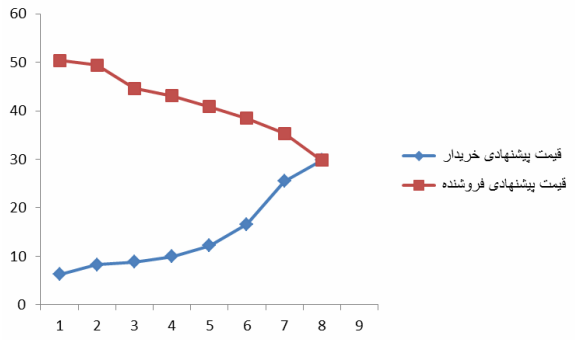
در شکل‌های ۵ و ۶ قیمت پیشنهادی هر دو روش بیان‌شده در طول مدت انجام یک مناقصه به‌صورت نمونه نشان داده شده است. همان طور که از شکل ۵ پیداست، قیمت توافقی در روش تجربی برابر ۲۹٫۷۷ است. در روش تجربی، خریدار هیچ گونه تلاشی را در راستای بالابردن سود خود انجام نمی‌دهد. در شکل ۶ معامله‌ای یکسان با استفاده از یادگیری تقویتی نشان داده شده است. قیمت توافقی در روش پیشنهادی ۲۴٫۹۵ است، لذا قیمت توافقی در روش پیشنهادی حدود ۱۶٪ نسبت به روش تجربی کاهش یافته است و به بیان دیگر سود حاصل از معامله برای خریدار به همین درصد افزایش یافته است. متقاضی برای کسب سود بیشتر زمان معامله را افزایش داده است.

٦- نتیجه‌گیری

در این مقاله به معرفی یک عامل هوشند برای چانه‌زنی در بازار پرداخته شد و چارچوب دو بازار مزایده و مناقصه شرح داده شد. سپس نحوه به‌کارگیری یادگیری تقویتی جهت یادگیری عامل شامل تعریف



شکل ۵: روند پیشنهاد قیمت در روش تجربی.



شکل ۶: روند پیشنهاد قیمت در روش پیشنهادی.

فضای حالت، انتخاب عمل و تعریف سیگنال تقویتی بیان شد. شبیه‌سازی انجام‌شده، کارایی بالای روش ارائه‌شده بر مبنای یادگیری تقویتی را از لحاظ به‌دست آوردن سود معامله و درصد پایین شکست‌ها در معامله نشان می‌دهد. نتایج شبیه‌سازی نشان داد که در شرایطی که فروشنده/ خریدار، وابسته به نوع مدل چانه‌زنی، طمع‌کار باشد، نمی‌تواند با طرف مقابل به توافق برسد و معامله به شکست منتهی می‌شود. استفاده از عامل هوشمند با انعطاف در ارائه پیشنهاد توانست بر این مشکل غلبه کند. ضمن آن که عامل هر گاه که شرایط اجازه دهد، سعی در بیشینه‌کردن سود خود به‌صورت حریصانه دارد. به‌طور مختصر می‌توان مزایای روش ارائه‌شده را به‌صورت زیر بر شمرد: ۱) چانه‌زنی هوشمند به‌منظور افزایش سود به وسیله تنظیم شیب منحنی پیشنهاد قیمت، ۲) حجم کم محاسبات و ۳) امکان یادگیری برخط.

همچنین به‌عنوان کار آینده می‌توان برای افزایش سرعت و کیفیت عملکرد، دانش اولیه در خصوص چانه‌زنی را با مقداردهی اولیه مناسب مقادیر ارزش- عمل در سیستم گنجاند.

مراجع

[1] A. Cakravastia and N. Nakamura, "Model for negotiating the price and due date for a single order with multiple suppliers in a make-to-order environment," *Int. J. of Production Research*, vol. 40, no. 14, pp. 3425-3440, 2002.

[2] A. Cakravastia and K. Takahashi, "Integrated model for supplier selection and negotiation in a make-to-order environment," *Int. J. of Production Research*, vol. 42, no. 21, pp. 4457-4474, 1 Nov. 2004.

[3] A. Cakravastia, I. S. Toha, and N. Nakamura, "A two-stage model for the design of supply chain networks," *Int. J. of Production Economics*, vol. 80, no. 3, pp. 231-248, 2002.

[4] N. Murthy, S. Soni, and S. Ghosh, "A framework for facilitating sourcing and allocation decisions for make-to-order items," *Decision Sciences*, vol. 35, no. 4, pp. 609-637, 2004.

[5] N. Sadeh and J. Sun, "Multi-attribute supply chain negotiation: coordinating reverse auctions subject to finite capacity considerations," in *Proc. of the 5th Int. Conf. on Electronic Commerce*, Pittsburgh, Pennsylvania, US, Oct. 2003.

Archive of SID
www.SID.ir

[6] Z. Ren and C. J. Anumba, "Learning in multi - agent systems: a case study of construction claims negotiation," *Advanced Engineering Informatics*, vol. 16, no. 4, pp. 265-275, 2002.
[7] D. Zeng and K. Sycara, "Bayesian learning in negotiation," *Int. J. of Human-Computer Studies*, vol. 48, no. 1, pp. 125-141, Jan. 1998.
[8] R. Carbonneau, G. E. Kersten, and R. Vahidov, "Predicting opponent's moves in electronic negotiations using neural networks," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1266-1273, Feb. 2008.
[9] H. Rau, M. H. Tsai, C. W. Chen, and W. J. Shiang, "Learning based automated negotiation between shipper and forwarder," *Computers & Industrial Engineering*, vol. 51, no. 3, pp. 464-481, 2006.
[10] K. L. Choy, W. B. Lee, H. C. W. Lau, D. Lu, and V. Lo, "Design of an intelligent supplier relationship management system for new product development," *Int. J. of Computer Integrated Manufacturing*, vol. 17, no. 8, pp. 692-715, Dec. 2004.
[11] S. Wei, J. Zhang, and Z. Li, "A supplier - selection using a neural network," in *Proc. IEEE Int. Conf. on Intelligent Proc. Systems*, pp. 468-471, Beijing, China, 1997.
[12] C. C. Lee and C. Ou-Yang, "A neural networks approach for forecasting the supplier's bid prices in supplier selection negotiation process," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2961-2970, Mar. 2009.
[13] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, Cambridge: MIT Press, 1998.
[14] C. Watkins and P. Dayan, "Q - learning," *Machine Learning*, vol. 8, no. 3, pp. 279-292, 1992.
[15] F. Saadatjoo, V. Derhmi, and V. Majd, "Balance of exploration and exploitation in deterministic and stochastic environment in Q-learning," in *Proc. 11st Int. Computer Conf.*, vol.1, pp. 492-498, Tehran, Iran, Jan.2006.

محمد علی سعادت‌جو در سال ۱۳۸۶ مدرک کارشناسی مهندسی کامپیوتر نرم‌افزار خود را از دانشگاه آزاد اسلامی میبد و در سال ۱۳۸۹ مدرک کارشناسی‌ارشد مهندسی کامپیوتر- هوش مصنوعی خود را از دانشگاه آزاد اسلامی قزوین دریافت نموده است. ایشان هم‌اکنون در مقطع دکتری مهندسی کامپیوتر- نرم‌افزار در دانشگاه کاشان مشغول به تحصیل است. زمینه‌های علمی مورد علاقه نام‌برده شامل تئوری صف، مهندسی نرم‌افزار، آموزش تقویتی، الگوریتم‌های تکاملی و شبکه‌های کامپیوتری می‌باشد.

ولی درهمی کارشناسی خود را در رشته مهندسی برق گرایش کنترل در سال ۱۳۷۵ در دانشگاه صنعتی اصفهان به پایان رساند. وی کارشناسی ارشد و دکترای خود را در رشته مهندسی برق-کنترل به‌ترتیب در سال‌های ۱۳۷۷و ۱۳۸۶ از دانشگاه تربیت مدرس دریافت کرد. دکتر درهمی از سال ۱۳۸۶ در گروه مهندسی کامپیوتر دانشگاه یزد مشغول به فعالیت گردید و اینک نیز عضو هیأت‌علمی این گروه می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان: یادگیری تقویتی، هوش مصنوعی گسترده، رباتیک، ناوبری ربات، فناوری اطلاعات، و یادگیری ماشینی می‌باشد.

فاطمه سعادت‌جو تحصیلات خود را در مقاطع کارشناسی، کارشناسی‌ارشد و دکتری در رشته ریاضیات کاربردی به‌ترتیب در سال‌های ۱۳۷۷، ۱۳۷۹ و ۱۳۸۸ در دانشگاه یزد به پایان رسانده است و هم‌اکنون استاد دانشکده کامپیوتر و معاون آموزشی موسسه آموزش عالی جهاد دانشگاهی یزد می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: آموزش تقویتی، پایگاه داده‌ها، الگوریتم‌های ژنتیک و کاربرد آنها در علوم.

Archive of SID

www.SID.ir