چکیده

چالش بزرگ مدیریت سفارش دهی در زنجیره تأمین اتخاذ یک راهبرد سفارش دهی مناسب در هر رده از زنجیره تأمین برای کاهش هزینه انبار است. در این پایان‌نامه عامل‌های هوشمند نرم‌افزاری وظیفه تصمیم‌گیری پویا را برای پیدا کردن یک راهبرد مناسب سفارش دهی بر عهده می‌گیرند. در گام اول ما به سراغ الگوریتم یادگیری Q[[1]](#footnote-1) می‌رویم و ضمن نشان دادن برتری آن نسبت به الگوریتم‌هایی که سیاست ثابت سفارش دهی را اتخاذ می‌کنند به بیان ضعف‌های آن در مدیریت سفارش دهی با فضای حالت بزرگ همچون سیستم‌های چند عامله می‌پردازیم .در گام بعد راهکاری را برای عمومیت بخشیدن به الگوریتم یادگیری Q با استفاده از تقریب زدن بیان می‌کنیم و نشان می‌دهیم که استفاده از الگوریتم بهینه‌شده با غلبه بر مشکل فضای حالت وسرعت یادگیری پایین، می‌تواند به خوبی روش‌های قبلی راهبرد بهینه را برای مدیریت سفارش دهی پیدا کند. این دست آورد سبب میشود که راه برای استفاده از این الگوریتم در مسایل دنیای واقعی که شامل شبکه ای از زنجیره های تامین با رده های مختلف سفارش دهی هستند هموار شود.

**کلمات کلیدی:** کنترل هوشمند موجودی، تعمیم یادگیری تقویتی، تخمین تابع ارزش، رگرسیون وزن دار محلی.

# مقدمه و مرور ادبیات

**زنجیره تأمین را می‌توان یک شبکه جهانی از تأمین‌کنندگان، کارخانجات، انبارها، مراکز توزیع و خرده‌فروشان دانست که موارد اولیه را گردآوری و تبدیل نموده و محصول نهایی را به دست مشتریان می‌رسانند. یک زنجیره تأمین شامل همه مراحل (اعضای زنجیره ) است که چه مستقیم و چه غیرمستقیم، در برآورده سازی درخواست یک مشتری نقش دارند. در یک زنجیره تأمین معمولی، مواد خام از انبارهای میانی و انبارهای توزیع‌کننده‌ها ارسال می‌شوند و از آنجا نیز به سمت خرده‌فروش‌ها و در نهایت به دست مشتری نهایی یا همان مصرف‌کننده می‌رسند. در بعضی از این مراحل، کالا انبارش می‌شود و در بعضی دیگر حمل می‌شود.** یعنی یک زنجیره تأمین مجموعه‌ای از انبارش ها و حمل‌ونقل ‌هاست**.**

**در جامعه‌ی صنعتی تولیدکنندگان کالا، یکی از مشکلات اساسی شرکت‌ها مدیریت موجودی انبار است. اگر موجودی انبار شرکتی بسیار زیاد باشد. شرکت می‌بایست، پیشاپیش هزینه‌ی تولید و استهلاک مازاد تولید را بپردازد. اگر موجودی انبار شرکت به هنگام افزایش تقاضای مصرف‌کننده بسیار کم باشد، شرکت فرصت فروش به مشتریانی را از دست خواهد داد که منتظر نمی مانند و ممکن است به سراغ رقیبی بروند که کالاهای مشابه دم دست دارد. هر چه مدیریت انبار بهتر باشد، سود بیشتری عاید شرکت می‌شود.کنترل موجودی همواره یک ابزار بسیار مهم و کارآمد برای شرکت‌ها است. به همین دلیل یکی از موضوعاتی که محققان همواره روی آن متمرکز داشته‌اند کنترل موجودی است. در گذشته مدیریت موجودی معمولا برای محصولات تمام‌شده در انتهای زنجیره به کار گرفته می‌شد. اما با گذشت زمان شرکت‌ها کم‌کم درک نمودند که** هزینه‌ی مازاد موجودی غیرضروری در هر کجای زنجیره تأمین که ذخیره باشد، خواه شرکت نسبت به آن مسئولیت مستقیم داشته باشد یا نه، باز هم در هزینه‌های پایین خطی شرکت تاثیر گذار است. **بنابراین امروزه موجودی نگهداری شده در هر یک از حلقه‌های زنجیره تأمین با یکدیگر مرتبط در نظر گرفته‌شده و به سطوح نگهداری مواد خام و اجزا با دید عناصر مرتبط به هم و به وجود آورنده‌ی فرصت برای کاهش هزینه‌ها نگریسته می‌شود.**

مدیریت سفارش در زنجیره تأمین که نقش مهمی در کنترل موجودی در زنجیره تأمین دارد عبارت است از یک رویکرد ترکیبی[[2]](#footnote-2) برای مشخص کردن مقدار سفارش هر عامل زنجیره تأمین که به عامل بالادست [[3]](#footnote-3) می‌دهد و هدف آن کمینه کردن هزینه انبار در کل زنجیره تأمین است.این رویکرد تمرکزش بر درخواست‌های زنجیره برای رسیدن به اهداف کاهش هزینه انبار کاهش کمبود کالا[[4]](#footnote-4) بالا بردن سطح سرویس به مشتری و افزایش سودمندی کل زنجیره است.در مواقعی که تقاضا از طرف مشتری مقدار ثابتی است سیاست بهینه در سفارش دهی "یک برای یک" یا "پاس دادن سفارش" است. بر اساس این سیاست هر عامل مقداری را سفارش میدهد که توسط پایین دستی به عنوان تقاضا رسیده است (S.O. Kimbrough, 2002)

**بیشتر کارهایی که جهت بهبود سفارش دهی در زنجیره تأمین انجام‌شده از الگوریتم ژنتیک برای بهینه‌سازی استفاده شده است . استفاده از الگوریتم ژنتیک دارای درجه آزادی بیشتری نسبت به سفارش دهی** یک بر یک30 **است. در این نوع بهینه‌سازی هر عامل برای کمینه کردن هزینه انبار در کل زنجیره تأمین با همکاری دیگر عامل‌ها قانون اختصاصی سیاست بهینه مخصوص به خود را یاد می‌گیرد .**

**یکی از محدودیت‌های الگوریتم‌های مبتنی بر ژنتیک قانون سفارش دهی ثابت****[[5]](#footnote-5) برای هر عامل است، به عنوان مثال در تحقیق** (S.O. Kimbrough, 2002) **با عنوان "آیا عامل هوشمند می‌تواند زنجیره تأمین را مدیریت کند؟" که یکی از اولین الگوریتم‌های پیشنهادی برای بهینه‌سازی میزان هزینه زنجیره تأمین است. پس از تکرارهای زیاد برای بهینه‌سازی با استفاده از ژنتیک ، قانون‌های ثابتی را برای سفارش دهی به عنوان قوانین بهینه پیشنهاد می‌کند، به این صورت که برای سه عامل قانونی به صورت مشخص می‌کند و اندازه درخواست پایین‌دستی است. درخواستی که به بالادستی سفارش داده می‌شود برابر است .** مقدار 𝒚 توسط الگوریتم ژنتیک مشخص می‌شود و **به عنوان یک مقدار ثابت همواره به درخواست پایین‌دستی اضافه می‌شود** و به عنوان سفارش به بالادستی ارسال می‌شود. تحقیق فوق نشان داده که استفاده از چنین راهبردی نتایج بهتری نسبت به راهبرد پاس دادن سفارش یا1-1دارد.

**قانون سفارش دهی ثابت**35 **برای هر عامل در الگوریتم‌های مبتنی بر ژنتیک از نقاط ضعف آن محسوب می‌شود به خصوص در محیط‌های پویا که توزیع آماری درخواست‌های مشتری به درستی مشخص نیست. یکی از اولین تلاش‌هایی که برای ایجاد قانون سفارش دهی به صورت پویا انجام‌شده مربوط به** (Giannoccaro & Pontrandolfo, 2002) **، که در آن از تکنیک میانگین پاداش نیمه مارکوف[[6]](#footnote-6) برای مسئله کنترل موجودی در زنجیره تأمین استفاده شده است.** (Chaharsooghi, Heydari, & Zegordi, 2008) **نیز از یادگیری تقویتی برای بهینه‌سازی مدل بازی نوشابه[[7]](#footnote-7) استفاده کرده است. مدل‌سازی** مسئله به این صورت عامل‌ها را قادر می‌سازد که قوانین مختلفی در حالت‌های مختلف انبار اتخاذ کنند.**[[8]](#footnote-8)**

**مدل‌سازی بر مبنای یادگیری تقویتی عامل‌ها را قادر می‌سازد از قوانین مختلفی در سفارش دهی کالا استفاده کنند که به نوبه‌ی خود این روش انعطاف‌پذیری بیشتری نسبت به بهینه‌سازی‌های مبتنی بر ژنتیک دارد. در تحقیق** (Fazel Zarandi, 2013)**که از یادگیری تقویتی فازی استفاده‌شده زنجیره تأمین را به صورت ترکیبی از چندین عامل هوشمند در نظر گرفته است که وظیفه این عامل‌ها سفارش دادن است و در آن هر عامل سعی میکند با بررسی کل زنجیره مقدار سفارش بهینه را بیابد . عامل‌ها بر اساس هدف مشترکی با یکدیگر تعامل دارند در تحقیق فوق آن‌ها یک سیستم خطی زنجیره تأمین را در نظر گرفته‌اند که در آن چهار عامل سفارش‌دهنده وجود دارد که هر کدام از آن‌ها مسئول سفارش دهی در یک قسمت خاصی از زنجیره است هدف اصلی عامل‌ها کاهش هزینه نهایی سیستم است که در بلندمدت محاسبه می‌شود. همچنین در این تحقیق عدم قطعیت نیز وجود دارد.**

**در مسایل دنیای واقعی با فضای حالت بسیار بزرگ وعدم قطعیت موجود در آن روش‌های سنتی یادگیری تقویتی معمولا نمایش ضعیفی از کارایی دارند. در دهه‌ی اخیر شیوه‌های مبتنی بر تجرید[[9]](#footnote-9) برای غلبه بر این موضوع مورد توجه قرار گرفته است. این شیوه‌ها به صورت طبیعی منجر به معماری کنترلی سلسله مراتبی و الگوریتم‌های یادگیری مرتبط با این ساختار می‌شوند. در یادگیری تقویتی سلسله مراتبی[[10]](#footnote-10) وظیفه اصلی به یک سری از زیر وظایف که عموماً برای حل ساده‌تر هستند فرموله بندی می‌شود. راه‌حل‌های سلسله مراتبی اغلب جواب‌های نزدیک به بهینه‌ای ارائه می‌دهند و نسبت به تکنیک‌های یادگیری تقویتی هزینه مناسب تری را در زمان اجرا، زمان یادگیری و فضای مورد نیاز برای حل مسایل محض دارند.** (Driessens & Dˇzeroski, 2002a)

**یک راه دیگر برای بهبود کارایی این الگوریتم تخمین زدن خط مشی[[11]](#footnote-11) (تابع ارزش[[12]](#footnote-12)) عامل در یادگیری تقویتی است.** در یادگیری تقویتی عمده تمرکز بر روی طراحی یک تابع تخمین زننده با قابلیت بالا در تخمین زدن و کارایی محاسباتی خوب است.(Fazel Zarandi, 2013)**.**

**استفاده از شبکه‌های عصبی مصنوعی[[13]](#footnote-13) راه حل دیگری برای نگهداری ارزش هر عمل به جای استفاده از جدول می‌باشد.**

**فضای حالت بزرگ دو چالش برای مسئله یادگیری تقویتی دارد یکی از آن‌ها فضای ذخیره‌سازی است به طوری که به ازای هر حالت، عمل[[14]](#footnote-14) بهینه باید ذخیره شود و دیگری عمومیت بخشیدن[[15]](#footnote-15) است؛ یعنی تمام این حالت‌ها طی فرایند یادگیری باید چندین بار تست شوند تا مقدار بهینه برای آن حالت به دست آید. به طوری که** تجربه کم ممکن است داده مورد نیاز را برای هر حالت فراهم نکند**[[16]](#footnote-16). این مسئله باعث می‌شود الگوریتم یادگیری تقویتی کارایی لازم را نداشته باشد (فضای حالت بزرگ و یادگیری کند). چالش دیگر سرعت یادگیری پایین این الگوریتم است چرا که باید ابتدا به طور کامل فضای حالت مساله را جستجو کند و راهکار بهینه را از آن استخراج کند.**

**با توجه به مطالب بیان‌شده حل مسئله معضل ابعاد**21 **بالا و همچنین سرعت بخشیدن به فرایند یادگیری این الگوریتم در مسئله کنترلِ موجودیِ چند عامله در یادگیری تقویتی می‌تواند هردو مشکل را حل کند. در این مقاله ما با ارایه روشی برای بهبود الگوریتم یادگیری تقویتی مسکله پیچیدگی فضای حالت را برای مساله کنترل موجودی در زنجیره تامین را برطرف میکنیم .**

ادامه این مقاله به این صورت سازماندهی شده که در بخش بعد یادگیری تقویتی و الگوریتم مورد استفاده بیان شده ،در بخش 3 نحوه مدل کردن مساله کنترل موجودی به یک مساله یادگیری تقویتی مبتنی برعامل بیان شده و در بخش 4 بررسی نتایج حاصل ازا این تحقیق آورده شده است. ودرپایان نتیجه گیری آورده شده است

1. QLearning- [↑](#footnote-ref-1)
2. integrated approach - [↑](#footnote-ref-2)
3. upstream actor - [↑](#footnote-ref-3)
4. lower slacks - [↑](#footnote-ref-4)
5. Fixed Rule Ordering - [↑](#footnote-ref-5)
6. Semi Markov Average Reward(SMART)- [↑](#footnote-ref-6)
7. beerGame (Sterman, 1989) یک شبیه سازی شناخته شده در زنجیره تامین است که بسیاری از علاقه مندان را همچون محققان دانشگاهی برای شبیه سازی سیستم های سفارش دهی به خود جذب کرده است. پارامترهای بهینه در سیاست سفارش دهی بازی نوشابه هنگامی که درخواست های مشتری افزایش می یابد در دو حالت متفاوت تجزیه وتحلیل می شود .(سیاست سفارش دهی ثابت برای همه عاملها وسیاست متفاوت). نشان داده می شود که کمترین هزینه برای زنجیره (در شرایط بازی نوشابه) زمانی بدست می آید که بازیکنان سیاست های متفاوتی داشته باشند (برخلاف سیاست یکسان برای همه بازیکنان) (Strozzi, Bosch, & Zaldívar, 2007)- [↑](#footnote-ref-7)
8. در‏2.4 خواهید دید که الگوریتم یادگیری تقویتی چگونه از تجربه فعلی برای سفارش دهی در آینده استفاده میکند.این روش به الگوریتم این امکان را میدهد که سیاست سفارش دهی پویا داشته باشد. - [↑](#footnote-ref-8)
9. abstraction - [↑](#footnote-ref-9)
10. Hierarchical Reinforcement Learning(HRL) - [↑](#footnote-ref-10)
11. - policy [↑](#footnote-ref-11)
12. - Value Function [↑](#footnote-ref-12)
13. - Artificial Neural Network [↑](#footnote-ref-13)
14. - State-Action [↑](#footnote-ref-14)
15. - Generalization [↑](#footnote-ref-15)
16. - علت این مساله این است که، برای اینکه الگوریتم نگاشت بهینه را پیدا کند لازم است که چندین بار در هر حالت قرار بگیرد(در این مساله به عنوان مثال چندین بار وضعیت انبار با مقدار کالای 10=X را مشاهده کند)تا بر اساس تجربه ای که کسب میکند مقدار مناسب سفارش دهی را پیداکند. عمده تمرکز این الگوریتم استفاده از تجربه برای یادگیری است. [↑](#footnote-ref-16)