Archive of SID

بسمه تعالی

بکارگیری "آتوماتای یادگیری تقویتی انتخابی" برای حل مسأله فروشنده دوره‌گرد

Application Of "Selective Action Reinforcement Learning Automata" (SARLA) In Solving TSP

داود ابراهیمی
دانشجوی کارشناسی ارشد مدیریت اجرائی
دانشگاه شهید باهنر کرمان – دانشکده مدیریت و اقتصاد
بخش مدیریت
davode@gmail.com

غلامعلی حیدری
دانشجوی کارشناسی ارشد کنترل
دانشگاه شهید باهنر کرمان – دانشکده فنی و مهندسی
بخش مهندسی برق
gheydari@yahoo.com

چکیده – روش هوشمند "آتوماتای یادگیری تقویتی انتخابی " یک روش ابتکاری جهت حل مسائل بهینه سازی است، که ایده آن از رفتار موجودات زنده هنگام یادگیری برگرفته شده است. این تکنیک برای حل انواع مسائل برنامه ریزی پویا و شبکه قابل استفاده می باشد. در این مقاله برای آزمودن کارایی الگوریتم، مسأله فروشنده دوره‌گرد مدلسازی و حل شده است و در نهایت ضریب کارایی روش نسبت به جواب بهینه به دست آمده است.

توضیح مدل – مدل مورد بررسی در این مقاله مسأله فروشنده دوره گرد نامتقارن (ATSP) می باشد که مثالهای متعدد آن به همراه جواب بهینه از سایت TSP Library اخذ شده و جوابهای به دست آمده از الگوریتم با جواب بهینه مقایسه گردیده‌اند.

توضیح روش – مهمترین فعالیت سیستم عصبی هر موجود زنده یادگیری است. یادگیری به دو نوع کلی قابل تقسیم است : یادگیری تحت نظارت و یادگیری تقویتی. در نوع اول موجود یادگیرنده توسط یک آموزش دهنده هدایت می‌شود تا بتواند دانش صحیح را نسبت به تعامل با محیط کسب کند. اما در نوع تقویتی موجود یادگیرنده به تنهایی با انجام آزمایش و با سعی و خطا دانش مورد نیاز خود را کسب میکند.

از لحاظ زیست شناسی سازوکار یادگیری تقویتی در طبیعت به این صورت بیان میگردد: هرگاه که یک موجود زنده با محیط پیرامون خویش ارتباط برقرار کند و در برابر یک سری شرایط معین A دست به انجام یک عمل معین A` زند، به خاطر زنجیره علت و معلول اثر عمل او از سوی محیط به خودش باز می گردد. چنانچه این اثر برگشتی مطابق میل او باشد، پیوندهای عصبی مرتبط با "عمل A` در شرایط A" تقویت می شوند و این موجب می گردد تا چنانچه او باز هم در شرایط مشابه A واقع شود احتمال سرزدن عمل A` از او بیشتر از اعمال دیگر باشد. به طور مشابه اگر اثر برگشتی از محیط خلاف میل آن موجود باشد پیوندهای عصبی مربوطه تضعیف شده و احتمال رخدادن آن عمل در آن شرایط کاهش می یابد. به تغییرات این چنینی در سیستم عصبی، یادگیری تقویتی گفته می‌شود.

www.SID.ir

Archive of SID

روش مورد استفاده در الگوریتم کامپیوتری بر این مبنا استوار است که ابتدا یک رشته غیر تکراری از شهر ها به طور تصادفی و با توزیع گسسته یکنواخت انتخاب می گردد سپس هزینه مسیر های متناظر با شهر های آن رشته بدست می آید در تکرار دوم رشته غیر تکراری دیگری از شهر ها تولید شده و هزینه آن محاسبه میگردد اگر هزینه کمتر از تکرار قبل باشد چگالی احتمال مربوط به انتخاب شدن مسیر های بین شهر های رشته دوم با یک درصد معین افزیش می یابد به همین ترتیب چنانچه هزینه مسیرهای بدست آمده از یک تکرار کمتر از هزینه های بدست آمده باشد چگالی احتمال مسیرهای بین شهر های بدست آمده در آن تکرار بیشتر میشود و این تابع چگالی جدید در انتخاب تصادفی شهر ها در تکرار بعدی بکار میرود. چنانچه هزینه در یک تکرار از مینیمم هزینه های در تکرارهای قبل کمتر باشد میزان افزایش دامنه چگالی متناظر با آن نیز کمتر میگردد تا جایی که اگر هزینه بدست آمده از یک تکرار از متوسط هزینه ها بیشتر باشد هیچ تغییری توابع چگالی رخ نمی دهد. طبیعتاً حداکثر افزایش در تابع چگالی وقتی رخ می دهد که هزینه بدست آمده کمتر از مینیمم هزینه های قبلی باشد.

با تکرار نمودن الگوریتم به تعداد کافی به تدریج توابع چگالی احتمال انتخاب شدن مسیرها روی یک سری مسیرهای خاص متمرکز میشوند. در این حالت الگوریتم به حالت پایدار خود رسیده و به مسیر بهینه یا نزدیک به بهینه همگرا شده است.

مزایا و معایب:

۱ –کارایی خوب روش برای حل مسائل با فضای غیر محدب
۲ –امکان حل طیف گسترده ای از مسائل پویای قطعی و احتمالی، پیوسته و گسسته
۳ –قابلیت توسعه برای مدلهای دارای متغیرهای تصمیم توأم
۴ –کاهش سرعت همگرایی در مسائل دارای فضای شدنی بزرگ

نتایج – نتایج به دست آمده از حل مدلهای ATSP به شرح زیر است:

مدل	جواب بهینه	میانگین جواب به روش SARLA	درصد خطا
Ft53	۶۹۰۵	۸۳۴۸	۲۱
Ft70	۳۸۶۷۳	۴۷۱۱۴	۲۲
P43	۵۶۲۰	۵۸۲۳	۳
Ftv35	۱۴۷۳	۱۵۵۳	۵
Br17	۳۹	۴۵	۱۵
Ry48p	۱۴۴۲۲	۱۵۵۱۶	۷
		میانگین خطا	۱۲

www.SID.ir

Archive of SID

مراجع :

[1] R S.SUTTON A G.BARTO "REINFORCEMENT LEARNING AN INTRODUTION" MIT PRESS, CAMBRIDGE, MA, 1998.

[2] M. N. Howell, G. P. Frost, T. J.Gordon and Q. H. Wu, "Continuous Action Reinforcement Learning Applied to Vehicle Suspension Control," Mechatronics, pp. 263-276, 1997.

[۳] غلامعلی حیدری، علی اکبر قره ویسی، مسعود رشیدی نژاد "طراحی بهینه کنترل PI در سیستم کنترل سرعت موتور القایی با روش هوشمند آتوماتای یادگیری تقویتی ترکیبی" اولین کنگره مشترک سیستمهای فازی و هوشمند ایران، دانشگاه فردوسی مشهد ۷ الی ۹ شهریور۱۳۸۶.

[8] F. S. HILLIER, G.J.LIEBERMAN "INDTRODUCTION TO OPERATION RESEARCH" MCGRAW-HILL HIGHER EDUCATION, 7th EDITION, 2001 .

[4] M. Kashki, A. Gharaveisi, and F. Kharaman, Application of CDCARLA Technique in Designing Takagi-Sugeno Fuzzy Logic Power System Stabilizer (PSS), PSOP32 PECon 2006

[5] Beigy, H. and Meybodi, M.R. A New Continuous Action-Set Learning Automaton for Function Optimization, ISCIS 2003, pp. 960-967, Springer-Verlag, Berlin Heidelberg,2003

[6] D. Ernst, M. Glavic and L. Wehenkel. "Power systems stability control : Reinforcement learning framework".IEEE Transactions on Power Systems, Feb.2004, Vol. 19, pages 427-435.

[7] Tadepalli P., DoKyeong Ok. "Model-based average reward reinforcement learning". Artificial Intelligence, pp. 177-224, 1998.

[8] D. L. APPLEGATE et al. "THE TRAVELING SALESMAN PROBLEM: A COMPUTATIONAL STUDY" PRINCETON SERIES IN APPLIED MATHEMATICS, 2007.

[9] XIANYI ZENG, ZEYI LIU. "A learning automata based algorithm for optimization of continuous complex functions", Information science, pp. 165-175. 2005.

www.SID.ir