

مدل یادگیری Q سلولی و کاربردهای آن

رضا رستگار محمدرضا میبیدی

آزمایشگاه محاسبات نرم
دانشکده مهندسی کامپیوتر و فناوری اطلاعات
دانشگاه صنعتی امیرکبیر
تهران ایران

چکیده

اتوماتای سلولی برای مدل کردن سیستم‌هایی مناسب است که قطعیت در تغییر حالات سیستم وجود داشته باشد. در حالیکه اغلب سیستم‌های واقعی پیچیده بوده و ویژگی نویزی بودن و عدم قطعیت و احتمالی بودن در آنها دیده می شود و به همین دلیل برای مدل کردن چنین سیستم هایی استفاده از اتوماتای سلولی با قوانین قطعی منطقی به نظر نمی رسد. در این مقاله از ترکیب یادگیری Q با اتوماتای سلولی مدل جدیدی به نام یادگیری Q سلولی¹ معرفی می‌گردد. این مدل جدید با استفاده از قابلیت های یادگیری Q مشکل نبود عدم قطعیت در تغییر حالات در اتوماتای سلولی را تا حدودی مرتفع می سازد. کاربرد این مدل ترکیبی در مساله تخصیص کانال در شبکه های سلولی مخابراتی مورد بررسی قرار میگیرد.

کلمات کلیدی: اتوماتای سلولی، یادگیری Q، یادگیری Q سلولی، شبکه سیار سلولی مخابراتی، تخصیص کانال، یادگیری

۱ - مقدمه

یک از ویژگیهای مشترک سیستم‌هایی که در آنها اجزای تشکیل دهنده در نواحی مکانی مختلف قرار داشته و یا به دلیل محدودیتهای ارتباطی، امکان انتقال تمام اطلاعات بین تمام اجزا وجود ندارد، توزیع شدگی و استفاده از سیستم کنترل کننده غیرمرکزی می باشد. به این ترتیب که در هر جز یا مجموعه اجزای این سیستم کنترل کننده هایی قرار دارند که به دلیل مشخصات سیستم تنها قادر به دریافت بخشی از اطلاعات مربوط به بخشهای مشخصی از سیستم می باشند و به همین دلیل ایجاد هماهنگی بین اجزای سیستم کاری دشوار است. از طرف دیگر در بسیاری از این سیستمها عدم قطعیت در اطلاعات و وجود نویزپذیری از ویژگیهای مشخص و تعیین کننده رفتار سیستم است که ایجاد هماهنگی بین اجزا سیستم را دشوار تر می سازد[1].

یادگیری می تواند به عنوان یک راه کار برای ایجاد تطبیق پذیری در اکثر سیستم‌هایی که دارای فرایندهای تصمیم گیری بر اساس عدم قطعیت و اطلاعات ناقص می باشند مورد استفاده گیرد. با استفاده از یادگیری در جایگاههای مناسب در سیستم، هر جز سازنده سیستم می تواند حتی با دریافت اطلاعات ناقص و غیر قطعی، به صورت تدریجی و بر اساس معیارهای تعریف شده در سیستم به استراتژی بهینه کنترلی مورد نیاز خود دست یابد. در یادگیری تقویتی که یکی از انواع مهم مدل‌های یادگیری می باشد، یک عامل یادگیرنده در طی فرایند یادگیری با

¹ Cellular Q-Learning

تعاملات² مکرر با محیط، به یک سیاست کنترل بهینه می‌رسد. کارایی این تعاملات با محیط بوسیله پیشینه (کمینه) بودن پاداشی (جریمه ای) که از محیط گرفته می‌شود، ارزیابی می‌گردد. از آنجاییکه این روش یادگیری در محیطی بالادرنگ انجام می‌گیرد، می‌توان آنرا همزمان با فعالیت محیط انجام داد که در این صورت با تمام رخدادهای پیش‌بینی نشده بصورت یک تجربه جدید برخورد می‌شود و می‌توان از آنها برای بهبود کیفیت یادگیری استفاده کرد. مزیت عمده یادگیری تقویتی نسبت به سایر روشهای یادگیری عدم نیاز به اطلاعات بجز سیگنال تقویتی از محیط میباشد.

یکی از مدل‌هایی که در شبیه سازی و یا مدل کردن سیستمها مورد استفاده قرار می‌گیرد، اتوماتای سلول است که اجزای آن به صورت مکانی توزیع شده اند و اطلاعات از طریق قوانین محلی حاکم بر سیستم به صورت جزئی بین اجزا رد و بدل می‌شوند [23][22]. در اتوماتای سلولی، فضا بصورت یک شبکه ای از سلولها تعریف می‌گردد، زمان بصورت گسسته پیش می‌رود و قوانین آن بصورت سرتاسری است که از طریق آن در هر مرحله هر سلول، وضعیت جدید خود را با در نظر گرفتن وضعیت همسایه‌های خود بدست می‌آورد. قانون اتوماتای سلولی، نحوه تاثیر پذیرفتن یک سلول از سلولهای همسایه خود را مشخص می‌کند. یک سلول، همسایه سلول دیگر گفته میشود اگر بتواند آن سلول را در یک مرحله و براساس قانون حاکم تحت تاثیر قرار دهد. ویژگی‌های اساسی اتوماتای سلولی، فضای گسسته، زمان گسسته، محدودیت تعداد وضعیتهای ممکن هر سلول، یکسان بودن تمام سلولها، قطعی بودن قوانین، وابستگی قانون در هر سلول به مقادیر سلولهای اطراف آن و وابستگی قانون به مقادیر تعداد محدودی از مراحل قبل همسایه ها و خود سلول می‌باشند. در اتوماتای سلولی همگام³ عمل بروز در آوردن سلولها به صورت همگام و در اتوماتای سلولی نا همگام⁴ عمل بروز در آوردن سلولها به صورت ناهمگام انجام می‌گیرد.

یکی از مشکلات اتوماتای سلولی تعیین فرم قطعی قوانین مورد نیاز برای یک کاربرد خاص است. اتوماتای سلولی برای مدل کردن سیستمهایی مناسب است که قطعیت در تغییر حالات سیستم وجود داشته باشد در صورتیکه اغلب سیستمها نویزی و دارای عدم قطعیت می‌باشند و وضع قوانین برای آنها به صورت قطعی، منطقی به نظر نمی‌رسد. روشهای متفاوتی برای حل این مشکل پیشنهاد شده است. یکی از این روشها احتمالاتی کردن قوانین می‌باشد اما مشکل این رهیافت، محاسبه این احتمالات برای سیستمهای ناشناخته می‌باشد. آیا اتوماتای سلولی قابلیت استخراج این احتمالات را دارد؟ در [۱] با معرفی اتوماتای یادگیر سلولی گامی در حل این مساله برداشته شده است. هر اتوماتای یادگیر سلولی، از یک اتوماتای سلولی تشکیل شده است که هر سلول آن به یک یا چند اتوماتای یادگیر مجهز می‌باشد که وضعیت این سلول را مشخص می‌سازد. مانند اتوماتای سلولی، قانون محلی در محیط حاکم است و این قانون تعیین می‌کند که آیا عمل انتخاب شده توسط یک اتوماتا در سلول باید پاداش داده شود و یا اینکه جریمه شود. دادن پاداش و یا جریمه منجر بروز در آوردن ساختار اتوماتای یادگیر سلولی بمنظور نیل به یک هدف مشخص می‌گردد [۱][۱۹][۲۰][۱۸][21].

در این مقاله مدل دیگری با نام "یادگیرنده Q سلولی" برای مشکل فوق‌الذکر پیشنهاد می‌شود. یادگیرنده Q سلولی یک اتوماتای سلولی است که هر سلول آن به یک یا چند یادگیرنده Q که یکی از انواع یادگیری تقویتی میباشد مجهز است که وضعیت سلول را مشخص میکند. مانند اتوماتای سلولی، قانون محلی در محیط حاکم است و این قانون تعیین می‌کند که آیا عمل انتخاب شده توسط یک یادگیرنده Q در سلول باید پاداش داده شود و یا اینکه جریمه شود. دادن پاداش و یا جریمه منجر بروز در آوردن ساختار یادگیرنده Q سلولی بمنظور نیل به یک هدف مشخص می‌گردد. برای نشان دادن کاربرد یادگیرنده Q سلولی، از آن برای حل مسئله تخصیص کانال در شبکه های سلولی مخابراتی استفاده میشود.

ادامه مقاله بدین صورت سازماندهی شده است. در ابتدا در بخش ۲ به بررسی اتوماتای سلولی می‌پردازیم. در بخش ۳ یادگیری Q شرح داده میشود. در بخش ۴ مدل یادگیرنده Q سلولی ارائه می‌شود. در بخش ۵، کاربرد یادگیرنده Q سلولی در حل مسئله تخصیص کانال در شبکه های سلولی مخابراتی مورد بررسی قرار می‌گیرد. بخش ۶ نتایج شبیه سازیها را ارائه میکند. بخش پایانی نتیجه گیری میباشد.

² Interaction

³ Synchronous Cellular Automata

⁴ Asynchronous Cellular Automata

۲- اتوماتای سلولی

اتوماتای سلولی شبکه‌ای سلولی است که هر سلول می‌تواند k حالت (وضعیت) داشته باشد. در هر سلول یک اتوماتا با حالات محدود^۵ قرار دارد. در حالت یک بعدی، هر سلول دو همسایه نزدیک به خود دارد. در این حالت، وضعیت سلول i در زمان $t+1$ یعنی $a_i^{(t+1)}$ مطابق فرمول زیر بدست می‌آید.

$$a_i^{(t+1)} = \phi(a_{i-1}^{(t)}, a_i^{(t)}, a_{i+1}^{(t)}) \quad (1)$$

تابع ϕ انون اتوماتای سلولی نامیده میشود. همسایگی در اتوماتای سلولی یک بعدی را می‌توان بگونه‌ای بسط داد که از دو همسایه بیشتر را نیز شامل شود. یعنی می‌توان شعاع r را برای همسایگی در نظر گرفت. البته معمولاً نزدیک‌ترین همسایه‌ها را در نظر می‌گیریم. همچنین سلولها در اتوماتای سلولی می‌توانند در شبکه‌ای با هر ابعادی قرار گیرند که متناسب با بعد، تعاریف مربوط به همسایگی و قانون تغییر می‌یابند. متداولترین اتوماتای سلولی، اتوماتای سلولی دو بعدی است. چند نوع همسایگی مهم در این نوع اتوماتای سلولی همسایگی مور^۶، ون نیومن^۷، کول^۸ و اسمیت^۹ می‌باشند. در همسایگی مور برای هر سلول مرکزی، هشت سلولی همسایه و در همسایگی ون نیومن، چهار سلولی همسایه در نظر گرفته می‌شود. سایر همسایگی‌ها در شکل ۱ نشان داده شده است. به طور رسمی اتوماتای سلولی D بعدی توسط چهارتایی $\langle Z^D, A, N, \Phi \rangle$ تعریف میشود [۳] بطوریکه $A = \{0, 1, \dots, k\}$ مجموعه متناهی از مقادیر مجاز که هر سلول اتوماتای سلولی میتواند اختیار کند، N یک زیر مجموعه متناهی مرتب از فضای Z^D است و همسایگی را تعریف می‌کند. و $\Phi = A^s \rightarrow A$ قانون اتوماتای سلولی است که s با توجه به نوع قانون و تعریف همسایگی متغیر می‌باشد.

قوانین در اتوماتای سلولی به سه گروه، عمومی^۹، جمعی و جمعی خارجی تقسیم می‌شوند که در ادامه شرح داده شده است:

قانون عمومی: در این قانون مقدار یک سلول در مرحله بعدی، به مقدار تک تک سلولهای همسایه در حالت فعلی وابسته است. برای مثال قانون $\phi(1, 0, 0, 0, 0) = 1$ با فرض مدل همسایگی ون نیومن به معنای آن است که در صورت ۱ بودن سلول مرکزی و صفر بودن بقیه سلولهای همسایه، حالت بعدی سلول مرکزی ۱ باشد.

قانون جمعی^{۱۰}: در این قانون مقدار یک سلول در مرحله بعدی، به تعداد سلولهای با حالت‌های مختلف بستگی دارد. در این نوع قانون برخلاف قانون عمومی، توجه ای به تک تک سلولها نمی‌شود.

$$\phi(0 \times n_1, \dots, k \times n_k) = i \quad 0 \leq i \leq k \quad (3)$$

که اعداد ۰ تا k مقادیر مجازی است که یک سلول میتواند اختیار کند. برای مثال قانون $\phi(0 \times 2, 1 \times 2) = 1$ با فرض مدل همسایگی ون نیومن به معنای آن است که در صورت صفر بودن ۲ تا از همسایه‌ها و ۱ بودن بقیه، حالت بعدی سلول ۱ خواهد بود.

قانون جمعی خارجی^{۱۱}: تفاوت این قانون با قوانین جمعی در این است که در تعیین حالت بعدی سلول، حالت فعلی سلول نیز موثر است.

$$\phi(j, 0 \times n_1, \dots, k \times n_k) = i \quad 0 \leq i, j \leq k \quad (4)$$

که اعداد ۰ تا k مقادیر مجازی است که یک سلول میتواند اختیار کند. برای مثال قانون $\phi(1, 0 \times 2, 1 \times 2) = 1$ با فرض مدل همسایگی ون نیومن به معنای آن است که در صورت ۱ بودن مقدار فعلی خود سلول، صفر بودن ۲ تا از همسایه‌ها و ۱ بودن بقیه، حالت بعدی سلول ۱ خواهد بود.

⁵ Finite State Automata

⁶ Moore

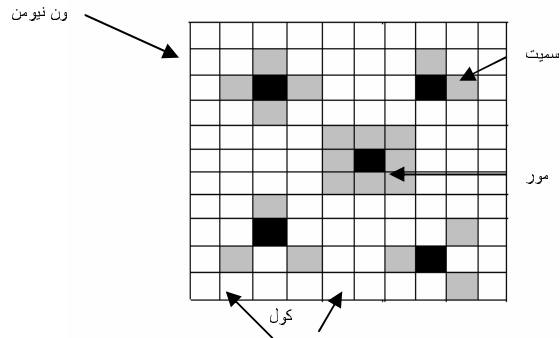
⁷ Cole

⁸ Smith

⁹ General

¹⁰ Totalistic

¹¹ Outer totalistic



شکل ۱: چهار نمونه از همسایگی متداول در اتوماتای سلولی دو بعدی (ون-نیومن، اسمیت، مور و کول)

۳- یادگیری Q

یک سیستم با مجموعه حالات محدود و قابل شمارش، S ، را در نظر بگیرید. یک کنترل کننده در هر حالت $s \in S$ ، یک عمل a را از میان مجموعه اعمال مجاز، $A(s)$ ، انتخاب و انجام می دهد. پس از آن سیستم با احتمال $p(s, a, s')$ ، از حالت s به s' می رود و پاداش $r(s, a, s')$ را به کنترل کننده می دهد. هدف کنترل کننده ماکزیمم کردن تابع زیر در تمامی حالات است:

$$J(s) = E\left\{\int_0^{\infty} e^{-\beta t} r(t) dt \mid s\right\} \quad (5)$$

که $E\{.\mid s\}$ امید ریاضی پاداش در یافت شده برای تمام مسیرهای ممکن آغاز شونده از s ، $r(t)$ نرخ عایدی کل در زمان t ، و β نرخ تخفیف^{۱۲} می باشند. در حالتی که زمان پیوسته نباشد می توانیم فرمول (۵) را به صورت زیر بنویسیم:

$$J(s) = E\left\{\sum_{t=0}^T \gamma^t r(t) dt \mid s\right\} \quad (6)$$

به طوری که γ نرخ تخفیف گسسته است.

احتمالهای انتقال حالت و پاداشها، تنها به حالت و عمل کنونی وابسته بوده و هیچ ارتباطی با گذشته ندارند. از این جهت این مدل یک مساله تصمیم گیری مارکوف^{۱۳} است و در حالتی که زمان گسسته باشد با نام مساله فرایندهای تصمیم گیری نیمه مارکوف^{۱۴} شناخته می شود. این مسائل را می توان با استفاده از برنامه سازی پویا^{۱۵} حل کرد. اما دو مانع بزرگ نفرین مدل^{۱۶} و نفرین بعد^{۱۷} مانع از این کار می شود. به همین دلیل الگوریتمهای یادگیری تقویتی بوجود آمدند. متدهای یادگیری تقویتی که بر پایه برنامه سازی پویا کار می کنند، از طریق تخمین های مناسب برای تابع مقدار بهینه شده^{۱۸}، J^* ، مساله فرایندهای تصمیم گیری نیمه مارکوف را حل می کنند. J^* طبقه معادله بلمن^{۱۹} به صورت زیر تعریف می شود:

$$J^*(s) = \max_{a \in A(s)} [E_{\Delta t, s'}\{c(s, a, s') + \gamma(\Delta t)J^*(s')\}] \quad (7)$$

که Δt زمان تصادفی تا رویداد بعدی، و $\gamma(\Delta t)$ نرخ تخفیف موثر برای حالت بعدی s' ، $\gamma(\Delta t) = e^{-\beta \Delta t}$ می باشند.

تخمین J^* می تواند از طریق یادگیری Q انجام گیرد. معادله بلمن را می توان با استفاده از مقادیر Q به صورت زیر بازنویسی کرد:

$$J^*(s) = \max_{a \in A(s)} Q^*(s, a) \quad (8)$$

¹² Discount Factor

¹³ Markov Decision Process

¹⁴ Semi-Markov Decision Process

¹⁵ Dynamic Programming

¹⁶ Curse of Modeling

¹⁷ Curse of Dimensionality

¹⁸ Optimal Value Function

¹⁹ Bellman Equation

در آغاز مقادیر به صورت تصادفی مقدار دهی می شوند و در ادامه با هر انتقال حالت در سیستم مقادیر Q بروز رسانی می شوند. و در حالت S عملی که بیشترین مقدار $Q(s,.)$ را داشته باشد انتخاب می گردد. اگر عامل یادگیرنده با انتخاب عمل a باعث انتقال سیستم به s' شود و پاداش $r(s, a, s')$ را دریافت کند، مقدار $Q(s, a)$ به صورت زیر تنظیم می شود:

$$Q(s, a) = (1 - \alpha)Q(s, a) + \alpha(r(s, a, s') + \gamma(\Delta t) \max_{b \in A(s')} Q^*(s', b)) \quad (9)$$

به طوری که $0 \leq \alpha \leq 1$ نرخ یادگیری می باشد. برای اطلاعات بیشتر در باره یادگیری Q میتوان به [۵ و ۱۶] مراجعه کنید.

۴- یادگیرنده Q سلولی

"یادگیرنده Q سلولی" که از این پس آن را با نام CQL می شناسیم مدلی است که از ترکیب اتوماتای سلولی و یادگیری Q حاصل میشود. این مدل برای سیستم های طراحی شده که اجزای آنها از طریق تعامل با یکدیگر از تجربیات گذشته همدیگر اطلاع پیدا میکنند و از این طریق می توانند رفتار خود را اصلاح کنند. یک CQL را می توان به صورت یک شش تایی $\langle E, A, N, \Phi, Q, C \rangle$ تعریف کرد که $E = \{e_1, \dots, e_n\}$ مجموعه مکانهای تعریف شده در اتوماتای سلولی هستند که می توانند در بشکلهای مختلفی مانند خطی، دو بعدی و سه بعدی در کنار هم قرار گیرند. $A = \{a_1, \dots, a_k\}$ مجموعه مقادیر مجازی است که یک سلول میتواند اختیار کند. $A^t(e_i)$ نشان دهنده مقدار سلول e_i در زمان t است. Φ قوانین حاکم بر جامعه سلولی است که پاداشها و جریمه ها براساس آن تعیین میشود. $N(e_i)$ مجموعه همسایه های سلول e_i را تعریف می کند که این مجموع دارای این ویژگیهاست:

$$e_i \notin N(e_i) \quad \forall e_i \in E \quad (10)$$

$$e_i \in N(e_j) \leftrightarrow e_j \in N(e_i) \quad \forall e_i, e_j \in E \quad (11)$$

هر قانون $\phi \in \Phi$ را می توان با توجه به مفاهیم قوانین عمومی و جمعی و یا وزن دار به یکی از فرمهای زیر تعریف کرد:

قوانین عمومی:

$$\langle a_1, \dots, a_h \rangle \rightarrow r \quad (12)$$

$$h = |N| + 1 \quad (13)$$

برای مثال قانون $1 \rightarrow \langle 1, 0, 0, 0, 0 \rangle$ با فرض همسایگی ون نیومن به معنای آن است که در صورت ۱ بودن سلول مرکزی و صفر بودن بقیه سلولهای همسایه، این خانواده جریمه شود. که این جریمه بر اساس تابع تخصیص پاداش C بین آنها تقسیم می شود.

قوانین جمعی:

$$\langle a_1 \times n_1, \dots, a_k \times n_k \rangle \rightarrow r \quad (14)$$

$$\sum_{i=1}^k n_i = |N| + 1 \quad (15)$$

برای مثال قانون $1 \rightarrow \langle 1, 0 \times 2, 1 \times 2 \rangle$ با فرض همسایگی ون نیومن به معنای آن است که در صورتی که مقدار سلول مرکزی ۱ باشد و از همسایه ها ۲ تا مقدار ۰ و ۲ تا مقدار ۱ باشند، این خانواده جریمه شود.

قوانین وزن دار:

$$\langle a_1 \times n_1, \dots, a_k \times n_k \rangle \rightarrow \sum_{i=1}^k n_i \times w_i \quad (16)$$

که $a_i \in A$ ، Γ مقادیر مجاز برای پاداش و تنبیه و w_i وزنی است که به هر وضعیت داده می شود. اگر $W(e_i) = N(e_i) \cup \{e_i\}$ ، در این صورت $C = C(W(e_i))$ ، $\phi = \phi(W(e_j))$. وظیفه تقسیم پاداش بدست آمده از ϕ مابین اعضای $W(e_i)$ را دارد. بر طبق تابع تخصیص پاداش C به هر سلول براساس کارایی آن پاداش داده میشود.

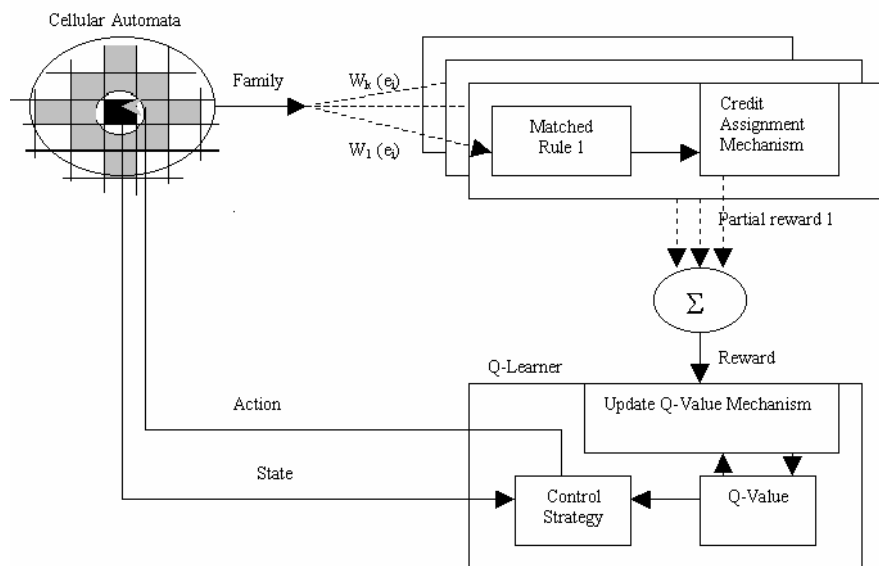
حال که با این مدل آشنا شدیم به شرح عملکرد آن می پردازیم. در نتیجه در یافت یک درخواست توسط سلول e_i ، این سلول فعال میشود (در مدل آسنکرون تمام سلولها به طور همزمان فعال نمی شوند) پس از فعال شدن، سلول، e_i ، از میان از مجموعه A یک مقدار انتخاب میکند

و سپس اعمالی بدین شرح را انجام می‌دهد. تابع ϕ برای سلول ارزیابی می‌شود و سپس طبق تابع C ، مقدار پاداش بدست آمده برای مجموعه $W(e_i)$ را بین اعضای آن تقسیم می‌کنیم. با چنین مکانیسمی در صورت سنکرون بودن اتوماتای سلولی، هر سلول e_i در گام t ام $|W(e_i)|$ خرده پاداش دریافت می‌کند که با جمع این خرده پاداش‌ها، پاداش کل سلول طبق (۱۸) بدست می‌آید که به یادگیرنده Q داده می‌شود و بر اساس آن مقادیر Q بهنگام می‌شوند.

$$sum = \sum_{\substack{\forall W(e_j) \\ \exists e_i \in W(e_j)}} C(W(e_j); e_i) \quad (18)$$

اگر اتوماتای سلولی به صورت آسنکرون عمل کند پس از دریافت سیگنال تقویتی، تنها سلول فعال شده و همسایه های آن بر طبق تابع تخصیص پاداش C پاداش را دریافت می‌کنند. سپس مقادیر پاداشها محاسبه شدند و به یادگیرنده‌های هر سلول داده می‌شود. نهایتاً سلول عمل بروز رسانی را انجام می‌دهد. مراحل فوق تا رسیدن سیستم به تعادل که در انجا مقدار هر سلول بهینه می‌باشد ادامه پیدا میکند.

همانطوریکه که گفته شد در مدل ارائه شده، پس از مشخص پاداش مربوط به یک $W(e_i)$ ، سیگنال پاداش بین سلولهای آن تقسیم می‌شود که این روش تقسیم توسط، $C(W(e_i))$ تعریف می‌شود و مقدار بازگشتی آن سهم پاداش هر سلول $e_j \in W(e_i)$ از مقدار $\phi(W(e_i))$ را نشان می‌دهد. روشهای متفاوتی برای پیاده‌سازی تابع تخصیص پاداش C وجود دارد که از مهمترین آنها می‌توان به تقسیم مساوی، تقسیم تصادفی، تقسیم براساس ویژگیهای سلول از جمله خبرگی [۲] سلول یا موقعیت جغرافیایی اشاره کرد.



شکل ۲: عملکرد یک سلول در مدل CQL

۵- حل مساله تخصیص کانال در شبکه های سیار سلولی توسط یادگیرنده Q سلولی

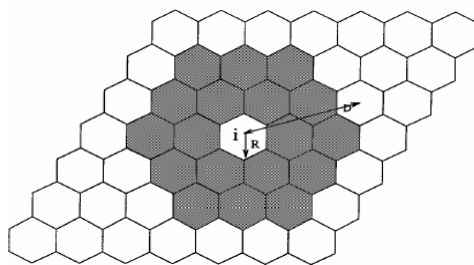
در این بخش مختصراً به مسئله تخصیص کانال در شبکه های سیار سلولی پرداخته می‌شود و سپس روشی مبتنی بر مدل یادگیرنده Q سلولی برای حل این مساله پیشنهاد میگردد.

۵-۱- تخصیص کانال در شبکه های سیار سلولی

در معماری سلولی، شبکه در حالت ایده آل متشکل از شش ضلعیهایی می باشد که نشان دهنده نواحی جغرافیایی می باشند (شکل ۳). درون سلولها، کاربران که میزبانهای سیار نامیده می شوند قادر به برقراری مکالمه می باشند. هر سلول دارای یک ایستگاه پایه است که کانالهای ارتباطی را در اختیار میزبانهای سیار قرار می دهد. سلولها به یک مرکز سوئیچینگ موبایل که مسئول کنترل مکالمات می باشد متصل می شوند. زمانی که یک کاربر موبایل در حال مکالمه به مرکز سلول می رسد، نیاز به تعویض کانال فعلی خود با کانال دیگری در سلول همسایه دارد. این نوع تغییر کانال در شبکه سلولی تحویل کانال^{۲۰} نامیده می شود. حال با فرض محدود بودن تعداد کانال ارتباطی و وجود محدودیت استفاده از یک کانال در دو سلول با فاصله ای کمتر از شعاع استفاده مجدد^{۲۱}، باید سعی شود از طریق اعمال استراتژی مناسب در اختصاص این کانالها به درخواستهای مکالمه به این اهداف دست یابیم. از جمله اصول مهمی که در معماری سلولی سبب دست یابی به اهداف فوق می گردد استفاده مجدد از کانال می باشد. استفاده مجدد کانال، استفاده از یک کانال در چندین سلول (سلولهای هم کانال) به طور همزمان می باشد.

استراتژیهای تخصیص کانال به دو گروه عمده تخصیص کانال ثابت^{۲۲} و تخصیص کانال دینامیک^{۲۳} تقسیم می شوند. در روش تخصیص کانال ثابت، تخصیص کانال برای هر سلول به طور ثابت و دائمی انجام می گیرد، به این صورت که بعد از طرحریزی دقیق فرکانس، به هر سلول یک مجموعه کانال اختصاص می یابد و این مجموعه ها برای مدت زمانی طولانی تغییر نخواهند کرد. ایستگاههای پایه در هر سلول فقط از کانالهای تخصیص یافته به سلول برای سرویس دهی به مکالمات درون سلول استفاده می کنند. استفاده از روش تخصیص کانال ثابت، ساده می باشد. ولی این روش به طور بهینه از کانال استفاده نمی کند. اگر بار ترافیک در سلول بیش از تعداد کانالهای نامی (کانالهایی که به طور ثابت به سلول اختصاص یافته اند) سلول باشد بار اضافی بدون توجه به شرایط در دیگر سلولها، بلوکه خواهد شد. این وضعیت برای شبکه های سلولی سیار بسیار محدود کننده می باشد و به همین دلیل استراتژی های مختلفی برای حداکثر کردن میزان استفاده از کانال و کم کردن نرخ بلوکه شدن مکالمات ارائه گردیده است. روشهایی که از استراتژی قرض گیری کانال استفاده می کنند، با اجازه دادن به یک سلول برای استفاده از کانالهای سلولهای دیگر، از مفهوم اولیه تخصیص کانال ثابت متمایز می شوند. قرض گیری کانال در یک سلول وقتی اتفاق می افتد که کانال آزادی در مجموعه کانال اختصاص یافته به سلول برای پشتیبانی از درخواست های کانال (که توسط مکالمات جدید یا تحویل کانال به وجود می آید) وجود نداشته باشد و یا تعداد کانال مشغول در یک سلول به یک حد آستانه خاص رسیده باشد [۶].

در استراتژی تخصیص کانال دینامیک هیچ گونه تخصیص کانال ثابت برای سلولهای شبکه وجود ندارد. تمام کانالهای فرکانسی سیستم در یک حوضچه مرکزی نگهداری می شوند. وقتی یک درخواست کانال در یکی از ایستگاههای پایه به وجود می آید، ایستگاه پایه با در نظر گرفتن تمام شرایط و محدودیت های تداخل سیگنال، مناسب ترین کانال که حداکثر کارایی را دارد انتخاب می کند. کانال انتخاب شده در طول مدت مکالمه در اختیار سلول می باشد و بعد از پایان یافتن مکالمه کانال به حوضچه مرکزی برگردانده می شود و یا برای استفاده در اختیار میزبان سیار دیگری در همین سلول قرار می گیرد [۱۱].



شکل ۳: ساختار شبکه سلولی مخابراتی، در این شبکه شعاع استفاده مجدد برابر سه می باشد.

²⁰ Handoff

²¹ Reuse Distance

²² Fixed Channel Assignment

²³ Dynamic Channel Assignment

مسئله تخصیص کانال یک مسئله NP-complete می باشد و بهمین دلیل الگوریتمهای تقریبی متعددی برای حل این مساله از جمله الگوریتمهای مبتنی بر الگوریتمهای ژنتیکی [۸]، تابکاری فلزات [۸]، جستجوی TABU [۸]، شبکه‌های عصبی [۹]، برنامه سازی پویای نوروئی [۴] و یادگیری Q [۱۰] گزارش شده اند.

در تمامی استراتژی‌های تخصیص کانال به صورت پویا، دانش و تجربه بدست آمده در طول کار سیستم به دست فراموشی سپرده می شود. اگر چه استراتژی‌های مبتنی بر شبکه‌های عصبی از آموزش بهره میبرند ولی در همه آنها داشتن یک ناظر خوب (یک استراتژی شناخته شده تخصیص پویای کانال) ضروری است. به نظر می رسد که استفاده از روشهای یادگیری بدون نیاز به داشتن استراتژی معین و شناخته شده بتواند گره گشای این مشکل باشد [۱۰]. در [۴ و ۱۰] دو استراتژی تخصیص کانال مبتنی بر الگوریتم های یادگیری تقویتی که در آنها نیازی به داشتن استراتژی معین و شناخته شده نیست ارائه گردیده است. در [۴] از یک معماری خطی تخمینی به همراه $TD(0)$ استفاده شده است. در این استراتژی هدف ماکزیمم کردن $E[\int_0^{\infty} e^{-\beta t} n(t) dt]$ می باشد که $n(t)$ تعداد مکالمات در جریان در زمان t و β نرخ تخفیف است. قابلیت توزیع شدگی و کارایی بالا از ویژگی های این الگوریتم است. در الگوریتم ارائه شده در [۱۱] از یادگیری Q برای تخصیص کانال به صورت تمرکز یافته^{۲۷} استفاده شده است. آنچه الگوریتم پیشنهادی در مقاله را از الگوریتم متمایز متمایز می سازد، مستقل شدن فرایند یادگیری در هر کدام از سلولهای شبکه می باشد که سبب می شود هر سلول بتواند بر اساس موقعیت جغرافیایی خود (به دلیل شرایط مرزی پوچ^{۲۸}، سیستم متقارن نمی باشد). استراتژی مناسب خود را یاد بگیرد. همچنین الگوریتم پیشنهادی کاملاً توزیع شده می باشد که این خود از حجم پیغامهای کنترلی بر روی شبکه میکاهد. الگوریتم پیشنهادی با دو الگوریتم فوقالذکر مقایسه خواهد شد.

۵-۲- تخصیص کانال با استفاده از یادگیرنده Q سلولی

در این قسمت به توصیف یک الگوریتم پویای توزیع شده تخصیص کانال مبتنی بر یادگیرنده Q سلولی اسنکرون می پردازیم. در الگوریتم پیشنهادی هر سلول در شبکه سلولی به یک سلول در یادگیرنده Q سلولی نگاشت می شود. زمانیکه یک سلول فعال شود (زمانیکه یک درخواست مکالمه به سلول میرسد)، یادگیرنده Q این سلول یک عمل انتخاب و سپس قانون حاکم بر یادگیرنده Q سلولی ارزیابی میشود. براساس نتیجه ارزیابی قانون، فرایند یادگیری انجام می شود. در ادامه این بخش پارامترهای یادگیرنده Q سلولی تعریف و سپس الگوریتم تخصیص کانال پویای مبتنی بر یادگیرنده Q سلولی ارائه میشود.

مدل همسایگی: مدل همسایگی، مدل همسایگی مور توسعه یافته^{۲۹} می باشد. مجموعه همسایه های هر سلول به دو دسته تقسیم می شوند: همسایه های ثابت و همسایه های متغیر. با فرض اینکه فاصله استفاده مجدد هم کانال R باشد، R حلقه حول سلول، سلولهای غیر همکانال یا همسایه های ثابت سلول می باشند که با سلول مرکزی تعامل محلی انجام می دهند. هر سلول در حالت کلی دارای $6(R+1)$ همسایه ثابت است که از این تعداد، حداکثر ۶ سلول همسایه مجاور می باشند و بقیه همسایگان غیر مجاور این سلول را تشکیل می دهند. هر سلول دارای یک مجموعه سلول هم کانال نیز می باشد که حالت آنها نیز در حالت فعلی سلول تاثیر دارد. مجموعه سلولهای همکانال یک سلول، سلولهایی هستند که در فاصله $R+1$ از سلول واقع هستند و می توانند از یک مجموعه کانال استفاده کنند. تعداد همسایه های متغیر یک سلول حداکثر ۶ تا می باشد. نکته ای که باید به آن توجه داشته باشیم آن است که با توجه به ویژگیهای فیزیکی شبکه مخابراتی سلولی اتوماتای سلولی استفاده شده در این الگوریتم دارای شرایط مرزی پوچ می باشد.

حالات سلول: در مدل استفاده شده هر سلول دارای متغیرهای حالت (U_i) متفاوتی است که تعداد آنها برابر تعداد کانالهای شبکه مخابراتی موبایل است و هر متغیر میتواند یکی از مقادیر *free* و *used* را اختیار کند.

قانون: قانون از نوع قانون وزن دار می باشد و تابعی از همسایه های متغیر و ثابت سلول می باشد. پاداش $r(s, a, s')$ پاسخ آنی سیستم در برابر تخصیص کانال a در حالت S می باشد که به صورت زیر تعریف میشود:

²⁴ Simulated Annealing

²⁵ TABU Search

²⁶ Neuro-Dynamic Programming

²⁷ Centralized

²⁸ Null Boundary Conditions

²⁹ Extended Moore

$$r(s, a, s') = n_1(k)w_1 + n_2(k)w_2 \quad (19)$$

که $n_1(k)$ تعداد سلولهای همکانال سلول i می باشد که در آنها از کانال K استفاده شده است و $n_2(k)$ تعداد سلولهای همکانال می باشد که در لایه سوم همسایگی (با شرط این که فاصله استفاده مجدد برابر دو باشد) قرار دارند و کانال K در آنها موجود است. ضرایب $w_1 = -1, w_2 = +1$ مقادیر ثابتی هستند.

تابع تخصیص پاداش: تابع تخصیص پاداش C به صورت زیر تعریف می شود:

$$C(W(e_i), e_j) = \begin{cases} 0 & \text{otherwise} \\ \phi(W(e_i)) = r(s, a, s') & e_j \text{ is central cell of } W(e_i) \end{cases} \quad (20)$$

یادگیرنده Q : در زیر پارامترهای یادگیرنده Q هر سلول e_i تعریف میشود.

- حالت: با فرض وجود N سلول و M کانال، حالت s به صورت زیر تعریف می شود (این تعریف مربوط به ساختار درونی یادگیرنده بوده و با تعریف حالات اتوماتای سلولی متفاوت است):

$$s = \sum_{i=1}^M H(i)2^{i-1} \quad (21)$$

به طوری که:

$$H(i) = \begin{cases} 1 & \text{Channel } i \text{ is not available} \\ 0 & \text{Otherwise} \end{cases} \quad (22)$$

- عمل: تخصیص یک کانال a از میان کانالهای $A(i)$ به درخواست رسیده به سلول i ام:

$$a = k, k \in A(i) \quad (23)$$

که $A(i)$ ، مجموعه کانالهای موجود^{۳۰} در سلول i است.

- حالت بعدی: با توجه به تعریف حالت که قبلا به آن اشاره شد، حالت بعدی از طریق واکنش الگوریتم به رویداد رسیده و حالت فعلی قابل محاسبه است.
- استراتژی کنترل: در اینجا از استراتژی کنترلی انتخاب Soft-Max^{۳۱} استفاده شده است که در آن احتمال انتخاب کانال k در حالت s بصورت زیر تعریف می شود:

$$e^{Q(s,k)/T} / \sum_{l=1}^M e^{Q(s,l)/T} \quad (24)$$

که T پارامتر دما است که با گذشت زمان کاهش می یابد. مقدار مینیمم T برابر ۱ میباشد.

- نرخ یادگیری: در الگوریتم پیاده سازی شده از نرخ یادگیری ثابت^{۳۲} و متغیر کاهش یابنده^{۳۳} استفاده شده است. نرخ یادگیری کاهش یابنده به صورت زیر تعریف می شود:

$$\alpha(s, a) = 1 / \text{visit}(s, a) \quad (25)$$

که $\text{visit}(s, a)$ تعداد دفعاتی است که عامل یادگیرنده، عمل a در حالت s را انتخاب می کند.

راههای متفاوتی برای ذخیره سازی مقادیر Q (شبکه عصبی، درخت تصمیم گیری^{۳۴} و جدول) وجود دارد. در این تحقیق از روش ذخیره سازی به شکل جدول استفاده شده است. از آنجا که حافظه مورد نیاز برای جدول، نمایی نسبت به تعداد کانالها در شبکه افزایش میابد و با

³⁰ Available

³¹ Soft-Max Selection Strategy Control

³² Constant Learning Rate

³³ Constant Learning Rate

³⁴ Decision Tree

توجه به اینکه در عمل تعداد کانالها زیاد است، حافظه موردنیاز برای ذخیره سازی مقادیر Q بسیار بالا خواهد بود. برای حل این مشکل فضای حالت را به چند مجموعه افراز کرده و برای هر مجموعه یک سطر در جدول Q ذخیره شده است.

حال به شرح الگوریتم تخصیص کانال پیشنهادی میپردازیم. هر گاه یک درخواست مکالمه به سلول i میرسد مراحل زیر انجام میگیرد.

۱- برای تمام سلولهای در همسایگی ثابت، پیام `give_used_channels` فرستاده می شود. سلولهای همسایه پس از دریافت این پیام با ارسال پیام `get_used_channels` که شامل لیست کانالهای اشغال شده خود می باشد به سلول i پاسخ می دهند.

۲- سلول i ، براساس پیغامهای دریافتی، حالت s را محاسبه می کند و از یادگیرنده Q خود بهترین کانال را درخواست می کند. یادگیرنده نیز براساس استراتژی کنترل `Soft-Max`، کانال k را انتخاب می کند. در صورتی که کانالی موجود نباشد، درخواست جدید مسدود می شود.

۳- پس از انتخاب کانال و انتساب آن، سلول i با ارسال پیام `lock_channel(k)` به تمام سلولهای در همسایگی ثابت خود از آنها می خواهد کانال k را در خود قفل کنند.

۴- پیام `give_channel_status(k)` به تمام همسایگان سلول متغیر ارسال می شود. هر همسایه پس از دریافت پیام، وضعیت کانال k را در پیام `get_channel_status(k)` به سلول i ارسال می کند.

۵- سلول i با آگاهی از وضعیت کانال k در تمام سلولهای همسایه مقدار پاداش خود را محاسبه و به یادگیرنده می دهد و یادگیرنده خود را به روز می کند.

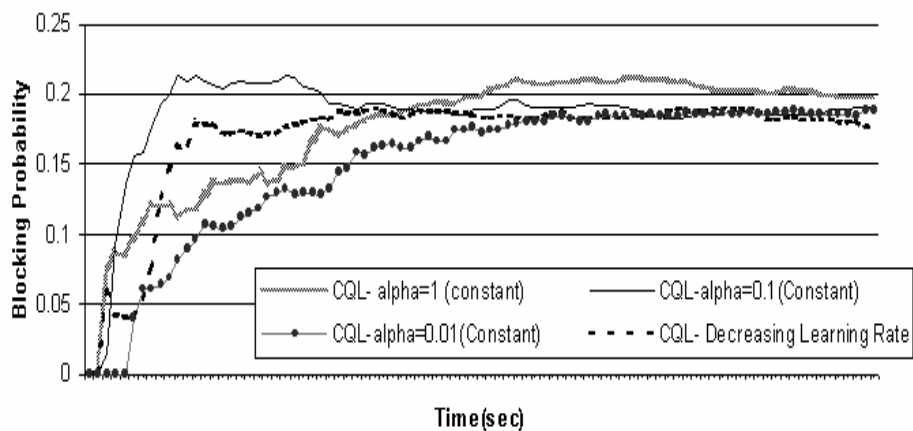
۶- پس از پایان مکالمه سلول i به تمام همسایگان ثابت خود پیام `unlock_channel(k)` ارسال می کند و از آنها می خواهد کانال k را آزاد کنند.

۶- نتایج شبیه سازی ها

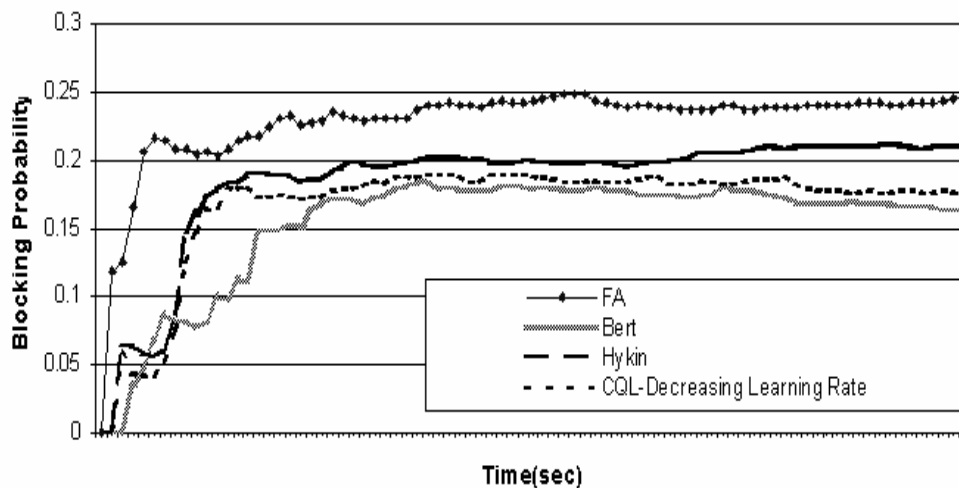
برای ارزیابی کارایی الگوریتم پیشنهادی (CQL-CA)، یک شبکه سلولی (اتوماتای سلولی) 7×7 در 7 با تعداد کانال 12 و شعاع استفاده مجدد 2 (12^2 حالت) در نظر گرفته شده است. متوسط زمان مکالمه و متوسط زمان تحویل کانال به ترتیب 3 و 2 دقیقه فرض شده است. در شبیه سازیها سه الگوریتم `FA`، `Bert` [۴] و `Hykin` [۱۱] پیاده سازی و با نتایج الگوریتم پیشنهادی مقایسه شده اند. همچنین کارایی الگوریتم پیشنهادی با نرخ یادگیری ثابت و متغیر کاهش یابنده مورد بررسی قرار دادیم. در تمام آزمایشها تحویل کانال در نظر گرفته شده است

در نمودار ۱، نمودار احتمال مسدود شدن مکالمات برای نرخ مکالمه 25 (مکالمه در ساعت) به ازای نرخهای یادگیری ثابت 0.1 ، 0.01 و 0.001 و نرخ یادگیری کاهش یابنده رسم شده است. همانطور که دیده می شود یادگیرنده با نرخ یادگیری کاهش یابنده در ابتدای شروع یادگیری کارایی نامناسبی از خود نشان می دهد اما با گذشت زمان میزان بهبود آن نسبت به میزان بهبود کارایی یادگیرنده های با نرخ یادگیری ثابت بسیار بهتر می باشد.

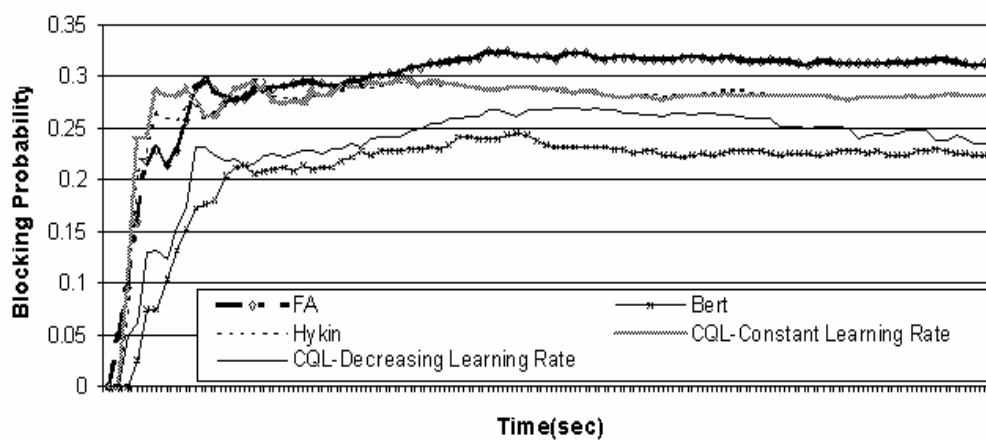
در نمودارهای ۲، ۳ و ۴، نمودار احتمال مسدود شدن مکالمات برای نرخهای مکالمه 20 ، 25 و 30 (مکالمه در ساعت) نشان داده شده است. همانطور که دیده می شود کارایی `CQL-CA` نسبت به کارایی الگوریتم `Hykin` بهتر بوده و در مقایسه با کارایی `Bert` این الگوریتم تعداد بیشتری از مکالمات را مسدود می کند. اما در عوض کاملاً توزیع شده می باشد و تعداد پیغامهای کنترلی کمتری بروی شبکه اعمال می کند.



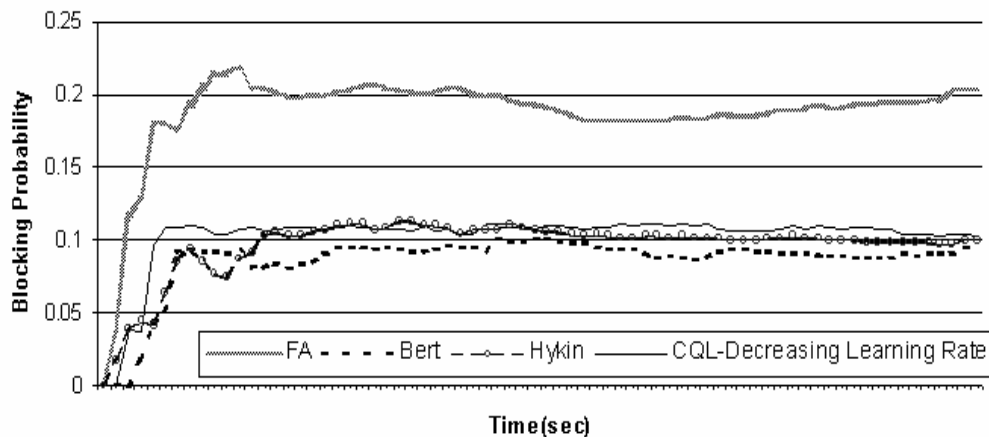
نمودار ۱: نمودار مقایسه نرخ مسدود شدن مکالمات برای الگوریتم CQL-CA برای نرخهای یادگیری ثابت مختلف و کاهش یابنده



نمودار ۲. نمودار مقایسه نرخ مسدود شدن مکالمات برای الگوریتمهای FA، Bert، Hykin و CQL-CA برای نرخ ترافیک متوسط



نمودار ۳. نمودار مقایسه نرخ مسدود شدن مکالمات برای الگوریتمهای FA، Bert، Hykin و CQL-CA برای نرخ ترافیک سنگین



نمودار ۴. نمودار مقایسه نرخ مسدود شدن مکالمات الگوریتمهای FA، Bert، Hykin و CQL-CA برای نرخ ترافیک سبک

۷- نتیجه گیری

در این مقاله یک مدل یادگیری جدید به نام یادگیرنده Q سلولی پیشنهاد گردید. یادگیرنده Q سلولی یک اتوماتای سلولی است که هر سلول آن به یک یا چند یادگیرنده Q مجهز می باشد که وضعیت این سلول را مشخص می سازد. مانند اتوماتای سلولی، قانون محلی در محیط حاکم است و تعیین می کند که آیا عمل انتخاب شده توسط یادگیرنده Q در یک سلول بایستی پاداش داده شود و یا اینکه جریمه شود. دادن پاداش و یا جریمه منجر به بروز در آوردن ساختار یادگیرنده Q سلولی بمنظور نیل به یک هدف مشخص میگردد. کاربرد این مدل ترکیبی برای حل مساله تخصیص کانال در شبکه های سلولی مخابراتی ارایه گردید. از طریق شبیه سازی نشان داده شد که الگوریتم پیشنهادی برای تخصیص کانال مبتنی بر این مدل ترکیبی از کارایی خوبی برخوردار است.

مراجع

- [1] Meybodi, M. R, Beigy, H. and Taherkhani, M., "Cellular Learning Automata", Proceedings of 6th Annual CSI Computer Conference, Computer Engineering Department, University of Isfahan, pp. 153 –163, 20-22 Feb. 2001.
- [2] Ahmadabadi, M. N. and Asadpour, M., "Expertness Based Cooperative Q-Learning", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 32, No. 1, 2002.
- [3] Adami, C., Introduction to Artificial Life, Springer Verlag, 1999.
- [4] Bertsekas, D. P. and Singh, S., "Reinforcement Learning for Dynamic Channel Allocation in Cellular Telephone Systems", NIPS96 Proceeding, 1996.
- [5] Bertsekas, D. P. and Tsitsiklis, J. N., Neuro-Dynamic Programming, Athena Scientific, Belmont, Massachusetts, 1996.
- [6] Brown, T. and Tong, H., "Adaptive Resource Allocation in Telecommunications", Denver, Proceeding of the SPIE, 1999.
- [7] Boukerche, A. and Jacob, T., "A Distributed Algorithm for Dynamic Channel Allocation", Kluwer Academic Publishers, Netherlands, Mobile Networks Journal, Vol. 7, PP. 115-126, 2002.
- [8] Chen, J., Seah, D. and Xu, W., "Channel Allocation for Cellular Networks Using Heuristic Methods", unpublished report, 1999.
- [9] Funabiki, N., "A Neural Network Parallel Algorithm for Channel Assignment Problems in Cellular Radio Networks", IEEE Transactions on Vehicular Technology, Vol 41, No. 4, 1992.
- [10] Hykin, S. and Nie, J., "A Dynamic Channel Assignment Policy through Q-learning", IEEE Transactions on Neural Networks, Vol. 10, No. 6, 1999.
- [11] Katzela, I. and Naghshineh, M., "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey", IEEE Personal Communications, 1996.
- [12] Krumke, S., Marathe, M. and Ravi, S., "Approximation Algorithms Assignment in Radio Networks", Dallas, International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications, 1996.

- [13] Lawrence, K. and Peter, Yum T., “Phantom Cell Analysis of Dynamic Channel Assignment in Cellular Mobile Systems”, IEEE Transactions on Vehicular Technology, Vol. 47, No. 1, 1998.
- [14] Lawrence Young, K. and Yum, K., “Compact Pattern Based Dynamic Channel Assignment for Cellular Mobile Systems”, IEEE Transactions on Vehicular Technology, Vol. 43, No. 4, 1994.
- [15] Smith K., “A Genetic Algorithm for the Channel Assignment Problem”, unpublished report, 1998.
- [16] Sutton, R., and Barto, A., Reinforcement Learning: An Introduction, MIT Press, 1998.
- [17] Tong, H. and Brown, T. X., “Reinforcement Learning for Call Admission Control and Routing under Quality of Service Constraints on Multimedia Networks”, Accepted in Machine Learning Journal, 2000.
- [18] Meybodi, M. R. and Kharazmi, M. R., “Cellular Learning Automata and Its Application to Image Processing”, Journal of Amirkabir, Vol. 14, No. 56A, pp. 1101-1126, 2004.
- [19] Meybodi, M. R, Beigy, H. and Taherkhani, M., “Cellular Learning Automata and Its Applications”, Journal of Science and Technology, University of Sharif, No. 25, pp.54-77, Autumn/Winter 2003-2004.
- [20] Beigy, H. and Meybodi, M. R., “A Mathematical Framework for Cellular Learning Automata”, Advances on Complex Systems, Vol. 7, Nos. 3-4, pp. 295-320, September/December 2004.
- [21] Beigy, H. and Meybodi, M. R., “Open Synchronous Cellular Learning Automata”, Journal of Computer Science and Engineering, ۲005, to appear.
- [22] Wolfram, S. Cellular Automata as Models of Complexity, Nature, 311, pp. 419—424, 1984.
- [23] Wolfram, S., Cellular Automata and Complexity, Perseus Books Group, 1994.