

حل مسئله حساب زبانی با استفاده از الگوریتم ترکیبی ژنتیک و اتوماتای سلولی احتمالی

امیر نورعلیعی^{*}، محمدرضا میبیدی[†]

چکیده

در این مقاله، یک الگوریتم ترکیبی مبتنی بر الگوریتم ژنتیک و اتوماتای سلولی احتمالی جهت حل مسئله حساب زبانی ارائه شده است. مسئله حساب زبانی یک معمای ریاضی و جزء مسائل NP-hard است. از الگوریتم ژنتیک بعنوان یک الگوریتم تکاملی و برای بهبود آن از اتوماتای سلولی احتمالی استفاده شده است. از قوانین احتمالی در اتوماتای سلولی احتمالی برای انتخاب والد ها در یک همسایگی هشت تایی و روشی جهت مهاجرت کروموزوم های با برازندگی بالا به همسایگی های مجاور استفاده شده است. برای ارزیابی الگوریتم ارائه شده آنرا بر روی پنج مسئله با پیچیدگی های متفاوت و استفاده از عملگر های مختلف آزمایش و نتایج را با الگوریتم ژنتیک و الگوریتم سلولی مقایسه کردیم. نتایج نشاندهنده بهبود کارایی الگوریتم است.

کلمات کلیدی

حساب زبانی، الگوریتم ژنتیک، اتوماتای سلولی احتمالی

^{*} دانشجوی کارشناسی ارشد هوش مصنوعی، دانشگاه آزاد اسلامی، واحد قزوین، دانشکده برق رایانه و فناوری اطلاعات، قزوین، ایران amir.nooraliei@qiau.ac.ir

[†] استاد تمام، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، تهران، ایران، mmeybodi@aut.ac.ir

1- مقدمه

حساب زبانی یک معمای ریاضی است که در آن ارقام بوسیله کاراکترها و یا نمادها جایگزین شده‌اند. حساب زبانی اولین بار توسط دودنی معرفی و در سال 1924 به چاپ رسید [1]. در این بازی مجموعه‌ای از معادلات وجود دارد که در آن بجای اعداد از حروف استفاده شده است. هر یک از معادلات می‌تواند شامل عملیات های ضرب، جمع، تفریق و تقسیم باشد. هدف یافتن مقدار متناظر هر حرف در معادله است. حساب زبانی جزء مسائل NP-hard می‌باشد [2].

از روشهای زیادی جهت حل مسائل NP-hard استفاده می‌شود. الگوریتم های تکاملی یکی از این روشها است. همان طور که تاریخ الگوریتم های تکاملی نشان می‌دهد، گونه های زیادی از الگوریتم‌های تکاملی وجود دارند [3]. ولی ایده همه آنها یکی است: با داشتن یک جمعیت اولیه، انتخاب بهترین های جمعیت و این افزایش شایستگی جمعیت را نتیجه می‌دهد. با داشتن یک تابع ارزیابی که می‌خواهیم بیشینه شود، می‌توان مجموعه ای از جواب های کاندید را به طور تصادفی تولید کرد و با تابع ارزیاب کیفیت آنها را به عنوان معیاری برای محاسبه شایستگی به کار برد بر اساس این شایستگی، بعضی از کاندیدهای بهتر انتخاب می‌شوند، تا به عنوان هسته ای برای تولید نسل بعد به کار روند. بر روی این کاندیدها عملگر های ترکیب و یا جهش اعمال می‌شود. عملگر ترکیب بر روی دو یا بیشتر کاندید اعمال می‌شود که والدین نام دارند و نتیجه آن تولید فرزند (فرزندانی) می‌شود.

اعمال ترکیب و جهش باعث تولید مجموعه جدیدی می‌شود که با مجموعه قبلی (والدین) رقابت می‌کنند تا در نهایت برنده ها در نسل بعدی ظاهر شوند. تا زمانی که یک کاندید با ویژگی های کافی (جواب) به دست بیاید و یا اینکه محدودیت‌هایی که از قبل برای مسئله تعریف کرده ایم، ارضا شوند این کار می‌تواند ادامه پیدا کند. یکی از مشهورترین الگوریتم های تکاملی، الگوریتم ژنتیک است.

در الگوریتم ژنتیک یک جمعیت از افراد طبق مطلوبیت آنها در محیط بقا می‌یابند. افرادی با قابلیت‌های برتر، شانس ازدواج و تولید مثل بیشتری را خواهند یافت. بنابراین بعد از چند نسل فرزندان با کارایی بهتر بوجود می‌آیند. در الگوریتم ژنتیک هر فرد از جمعیت بصورت یک کروموزوم معرفی می‌شود. کروموزومها در طول چندین نسل کاملتر می‌شوند. در هر نسل کروموزومها ارزیابی می‌شوند و متناسب با ارزش خود امکان بقا و تکثیر می‌یابند. تولید نسل در بحث الگوریتم ژنتیک با عملگرهای تقاطع و جهش صورت می‌گیرد. والدین برتر بر اساس یک تابع برازندگی انتخاب می‌شوند. الگوریتم‌های ژنتیکی را می‌توان یک روش بهینه‌سازی تصادفی جهت‌دار دانست که به تدریج به سمت نقطه بهینه حرکت می‌کند [4].

اتوماتای سلوای در اواخر دهه 40 توسط جان وان نیومن مطرح شد و پس از او توسط ریاضی دانی بنام استانیسلاو اولام به عنوان مدلی برای

بررسی رفتار سیستم های پیچیده پیشنهاد شد [5, 6]. اتوماتای سلولی سیستم‌های دینامیکی گسسته‌ای هستند که رفتارشان کاملاً بر اساس ارتباط محلی استوار است و می‌توانند برای محاسبات و شبیه سازی سیستم‌های دینامیکی به کار روند. فیزیک اولین علمی است که سیستم‌های دینامیکی را بررسی کرده است. اتوماتاهای سلولی با قوانین ساده و محلی می‌توانند محاسبات و رفتار پیچیده ای از خود بروز دهند. انواع متفاوتی از قوانین به روزرسانی سلولها وجود دارند که باعث ایجاد انواع متفاوت اتوماتای سلولی می‌شوند. قوانین می‌توانند به صورت قطعی و یا احتمالی بیان گردند و این دو دسته از قوانین منجر به دو دسته اتوماتای سلولی قطعی و اتوماتای سلولی احتمالی می‌شوند [7].

در این مقاله ما از ترکیب اتوماتای سلولی احتمالی و الگوریتم ژنتیک استفاده کردیم. از خاصیت احتمالی بودن اتوماتای سلولی احتمالی جهت انتخاب والدین در مرحله تقاطع بهره گرفته شده است. در اتوماتای سلولی از همسایگی مور استفاده شده است.

سازمان و ترتیب مقاله بصورت زیر است: مسئله حساب زبانی در بخش 2 توضیح داده شده است. در بخش 3، الگوریتم ژنتیک تشریح شده است. اتوماتای سلولی احتمالی در بخش 4 توضیح داده شده است. در بخش 5 الگوریتم ترکیبی پیشنهادی تشریح شده است. در بخش 6 نتایج آزمایشات نشان داده شده است و در بخش آخر نتیجه‌گیری توضیح داده شده است.

2- مسئله حساب زبانی

حساب زبانی یک بازی از زیر مجموعه های مسائل ارضای قیود است که جزء مسائل NP-hard می‌باشد. در این بازی مجموعه‌ای از معادلات وجود دارد که اعداد به صورت حروف نمایش داده می‌شوند. هر یک از معادلات می‌تواند شامل عملیات های ضرب، جمع، تفریق و تقسیم باشد. هدف یافتن مقدار متناظر هر حرف در معادله است. حداکثر تعداد حروف برای هر مسئله می‌تواند 10 حرف مختلف باشد که می‌تواند تکراری هم باشد. برای هر مسئله ممکن است چند جواب وجود داشته باشد. شکل 1 یک نمونه مسئله حساب زبانی را نشان می‌دهد.

$ \begin{array}{r} \text{S E N D} \\ + \text{M O R E} \\ \hline \text{M O N E Y} \end{array} $
شکل (1): نمونه مسئله حساب زبانی

جواب مسئله برابر مقادیر حروف $O = 0, M = 1, Y = 2, E = 5, N = 6, D = 7, R = 8, S = 1$ است. در حالت معمولی تعداد حالت های کلی برای جستجو برابر $8 \times 8!$ یا 322560 حالت است که

جستجوی این فضا نیاز به زمان زیادی دارد بنابراین برای حل آن از الگوریتم های تکاملی استفاده می کنند که یک از این الگوریتم ها الگوریتم ژنتیک است [8, 9].

3- الگوریتم ژنتیک

الگوریتم ژنتیک، الهامی از علم ژنتیک و نظریه تکامل داروین است. الگوریتم ژنتیک یک روش جستجوی احتمالی است که از شبیه سازی تکامل زیستی و طبیعی استفاده می کند. الگوریتم های ژنتیک با به کارگیری اصل بقای بهترین ها برای تولید تخمین های هر چه بهتر یک جواب روی جمعیتی از جواب های بالقوه عمل می نماید. در هر نسل، مجموعه ای از تخمین ها توسط فرایند انتخاب افراد مطابق با سطح برازندگی شان در دامنه مساله، و پرورش آنها با هم با استفاده از عملگرهای گرفته شده از ژنتیک طبیعی ایجاد می گردد. این فرایند، ما را به سمت تکامل جمعیت هایی از افراد، که با محیط مربوطه شان بهتر از والدین شان وفق داده شده اند هدایت می کند. یکی از مهمترین بخش های الگوریتم ژنتیک ساخت کروموزوم های اولیه با توجه به مسئله است. شکل 2 نحوه ساخت کروموزوم برای مسئله حساب زبانی را نشان می دهد. بدلیل اینکه هر کاراکتر معادل یک عدد است بنابر این طول کروموزوم ها باید 10 در نظر گرفته شود و هیچ کاراکتری نمی تواند معادل دو عدد یکسان باشد.

S	E	N	D	M	O	R	Y	-	-
1	2	3	4	5	6	7	8	9	0

شکل (2): نمایش یک کروموزوم

نرخ تقاطع 80% و از تقاطع جایگشتی استفاده شده است. نرخ جهش نیز 30% و از عملگر جهش جایگشتی استفاده شده است.

4- اتوماتای سلولی احتمالی

برای حل مسئله های زیادی از آتاماتای سلولی استفاده شده است. اتوماتای سلولی مدل ریاضی برای سیستم هایی است که در آنها چندین مؤلفه ساده برای الگوهای پیچیده با هم همکاری می کنند. اتوماتای سلولی از یک شبکه منظم سلولها تشکیل شده است که هر سلول می تواند ($K > 1$) مقدار مختلف به خود بگیرد. سلولهای اتوماتای سلولی در زمانهای گسسته بطور همزمان و برطبق یک قانون محلی بنام Φ بهنگام می شوند که در آن مقدار هر سلول براساس مقادیر سلول های همسایه تعیین می گردد [10].

هر اتوماتای سلولی را بصورت 7 تایی به صورت $(Q, d, V, \Sigma, \Delta, \delta, \lambda)$ می باشد که در آن:

1- Q - مجموعه حالاتی است که هر سلول می تواند اختیار کند.

2- d ابعاد فضای سلولی را مشخص می نماید. به ازای $d=1$ یک اتوماتای سلولی یک بعدی و به ازای $d=2$ یک اتوماتای سلولی دوبعدی خواهیم داشت.

3- برای هر سلول x در اتوماتای سلولی آرایه $V(x) = \{x + v_0, x + v_1, \dots, x + v_k\}$ مشخص کننده $k+1$

همسایه ای می باشد که بصورت مستقیم با سلول در ارتباطند.

4- Σ الفبای ورودی اتوماتای سلولی می باشد.

5- Δ الفبای خروجی اتوماتای سلولی می باشد.

6- δ تابع انتقال است که بفرم $d : (Q \times \Sigma^n)^{k+1} \rightarrow Q$ می باشد.

براساس تابع انتقال، حالت بعدی هر سلول به حالت و مقادیر حافظه ای تمامی همسایگان آن سلول در مرحله فعلی بستگی دارد. تعداد ثبات های هر سلول است.

7- λ رابطه مبدل است. این مبدل مقدار هر حافظه سلول را با توجه به حالت و مقادیر حافظه ای همسایگان مشخص می سازد [11].

انواع متفاوتی از قوانین به روزرسانی سلولها وجود دارند که باعث ایجاد انواع متفاوت اتوماتای سلولی می شوند. قوانین میتوانند به صورت قطعی و یا احتمالی بیان گردند و این دو دسته از قوانین منجر به دو دسته اتوماتای سلولی قطعی و اتوماتای سلولی احتمالی می شوند [7]. استفاده از قوانین احتمالی در اتوماتای سلولی باعث افزایش قدرت اتوماتای سلولی در شبیه سازی پدیده های طبیعی می شود. در اتوماتای قطعی حالت یک سلول به طور قطعی توسط حالت فعلی و حالات همسایگان تعیین می شود در حالیکه در اتوماتای سلولی احتمالی قوانین احتمالی است و هر قانون احتمال مربوط به خود را دارد ممکن است در یک قانون احتمالی یک موقعیت یکسان خروجی متفاوتی تولید نماید [12].

5- الگوریتم پیشنهادی

در این مقاله ما الگوریتم ژنتیک را با اتوماتای سلولی احتمالی ترکیب کردیم. مراحل الگوریتم پیشنهادی به صورت ذیل است.

ابتدا به هر سلول از اتوماتای سلولی یک کروموزوم اختصاص داده می شود سپس مقدار برازندگی هر کروموزوم محاسبه شده و یک مقدار احتمال برای هر سلول در نظر گرفته می شود که این مقدار با مقدار برازندگی آن سلول متناسب است و هرچه سلول برازندگی بیشتری داشته باشد مقدار احتمال اختصاص داده شده به آن هم بیشتر می شود. در مرحله بعد با استفاده از همسایگی مور در اتوماتای سلولی احتمالی عمل انتخاب از بین هشت سلول همسایه سلول مرکزی انتخاب می شود در این مرحله با استفاده از قوانین احتمالی اتوماتای سلولی احتمالی هر سلول با توجه به قانون اتوماتا و مقدار احتمال خود انتخاب می شود در این حالت حداقل دو و حداکثر هر هشت همسایه برای مرحله تقاطع انتخاب می شوند. در مرحله تقاطع و جهش با استفاده از عملگر تقاطع و جهش جایگشتی عمل تقاطع و جهش انجام شده و

سپس کروموزوم با بالاترین برازندگی انتخاب می شود و به جای کروموزوم سلول مرکزی قرار داده می شود و این کار تا زمانی که به جواب برسیم برای تمامی سلولها ادامه پیدا می کند. با استفاده از یک مقدار تصادفی به عنوان احتمال مهاجرت در هر بار اجرا سلول مرکزی به یکی از هشت همسایه خود مهاجرت می کند.

6- نتایج آزمایشات

برای آزمایش کارایی الگوریتم پیشنهادی، ما آنرا بر روی پنج مسئله طراحی شده با تعداد کاراکتر ها و عملگر های متفاوت آزمایش کردیم. این مسائل با توجه به تعداد کاراکترهای موجود در آنها از خیلی ساده تا خیلی سخت طبقه بندی شده اند. جدول 1 مسائل طراحی شده را نشان می دهد.

جدول (1): مسائل حساب زبانی

شماره مسئله	نوع مسئله	تعداد حروف	نمایش مسئله
1	خیلی ساده	3	$\begin{array}{r} N \\ + M \\ \hline S \end{array}$
2	ساده	6	$\begin{array}{r} T W O \\ + T W O \\ \hline F O U R \end{array}$
3	متوسط	7	$\begin{array}{r} B A S E \\ + B A L L \\ \hline G A M E S \end{array}$
4	سخت	8	$\begin{array}{r} S E N D \\ + M O R E \\ \hline M O N E Y \end{array}$
5	خیلی سخت	10	$\begin{array}{r} C R O S S \\ + R O A D S \\ \hline D A N G E R \end{array}$

معیار های تعداد رسیدن به جواب، مجموع اختلافات، کمترین و بیشترین اختلاف با جواب الگوریتم پیشنهادی را در تعداد اجراهای 50 و 100 و جمعیت های اولیه 16، 64 و 100 با الگوریتم ژنتیک و الگوریتم ژنتیک سلولی مقایسه کردیم. نتایج از 15 بار اجرای الگوریتم و استفاده از عملگر جمع بدست آمده است.

پارامتر						مقدار
جمعیت اولیه						16
تعداد اجرا						50
شماره مسئله	الگوریتم	1	2	3	4	5
تعداد رسیدن به جواب	GA	12	3	0	0	0
	CA-GA	15	11	0	2	0
	PCA-GA	15	12	1	5	0
کمترین اختلاف با جواب	GA	0	0	1	2	10
	CA-GA	0	0	1	0	3
	PCA-GA	0	0	0	0	2
بیشترین اختلاف با جواب	GA	1	15	204	17	922
	CA-GA	0	1	21	8	140
	PCA-GA	0	2	23	17	106
مجموع اختلافات با جواب	GA	3	43	413	43	3802
	CA-GA	0	4	68	39	494
	PCA-GA	0	4	123	50	424

(الف)

پارامتر						مقدار
جمعیت اولیه						16
تعداد اجرا						100
شماره مسئله	الگوریتم	1	2	3	4	5
تعداد رسیدن به جواب	GA	15	6	0	4	0
	CA-GA	15	13	1	5	0
	PCA-GA	15	15	2	6	0
کمترین اختلاف با جواب	GA	0	0	1	0	6
	CA-GA	0	0	0	0	1
	PCA-GA	0	0	0	0	1
بیشترین اختلاف با جواب	GA	0	3	30	13	296
	CA-GA	0	1	5	7	108
	PCA-GA	0	0	4	4	48
مجموع اختلافات با جواب	GA	0	14	99	26	1120
	CA-GA	0	2	28	27	268
	PCA-GA	0	0	29	19	257

(ب)

پارامتر						مقدار
جمعیت اولیه						64
تعداد اجرا						50
شماره مسئله	الگوریتم	1	2	3	4	5
تعداد رسیدن به جواب	GA	15	12	1	8	0
	CA-GA	15	15	3	7	0
	PCA-GA	15	15	3	9	0
کمترین اختلاف با جواب	GA	0	0	0	0	2
	CA-GA	0	0	0	0	1
	PCA-GA	0	0	0	0	1
بیشترین اختلاف با جواب	GA	0	3	3	10	44
	CA-GA	0	0	3	9	18
	PCA-GA	0	0	3	8	24
مجموع اختلافات با جواب	GA	0	5	19	24	286
	CA-GA	0	0	17	20	120
	PCA-GA	0	0	15	20	105

(ج)

زیر نویس‌ها

[1] H. E. Dudeney, in Strand Magazine, vol. 68, pp. -97 214, 1924

[2] H. Soni and N. Arora, "SOLVING CRYPT-ARITHMETIC PROBLEMS VIA GENETIC ALGORITHM," rin, vol. 2229, p. 6115

[3] A. E. Eiben and J. E. Smith, Introduction to evolutionary computing: springer, 2008

[4] L. Davis, Handbook of genetic algorithms vol. 115: Van Nostrand Reinhold New York, 1991

[5] J. Von Neumann, "The general and logical theory of automata," Cerebral mechanisms in behavior, pp. 41-1, 1951

[6] J. Von Neumann and A. W. Burks" ,Theory of self-reproducing automata," 1966

[7] A. RAHIMI, et al., "A PROBABILISTIC CELLULAR AUTOMATA BASED ALGORITHM FOR LOAD BALANCING IN CELLULAR MOBILE SYSTEMS," DANESHVAR MEDICINE, 2003

[8] M. S. K. Namdeo, et al., "Crypto-Arithmetic Problem using Parallel Genetic Algorithm (PGA)".(

[9] N. Shedge Kishor, et al., "Solving Verbal Crypto-Arithmetic Problem by Parallel Genetic Algorithm (PGA)," International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume, vol. 2

[10] A. A. Abin, et al., "A new dynamic cellular learning automata-based skin detector," Multimedia systems, vol. 15, pp. 323-309, 2009

[11] م. شاه آبادی م. ر. میبیدی، "پایه سازی الگوریتم انتشار نور

توسط اتوماتای سلولی دو بعدی،" اولین کنفرانس بین المللی فناوری

اطلاعات و دانش، تهران، دانشگاه صنعتی امیرکبیر، 1382، 48-36pp.

[12] N. Boccara and H. Fuks, "Modeling diffusion of innovations with probabilistic cellular automata," Cellular Automata: A Parallel Model, pp. 280-263, 1998

پارامتر						مقدار	
جمعیت اولیه تعداد اجرا						100	
						100	
شماره مسئله	الگوریتم	1	2	3	4	5	
تعداد رسیدن به جواب	GA	15	14	0	11	0	
	CA-GA	15	15	13	14	0	
	PCA-GA	15	15	13	14	1	
کمترین اختلاف با جواب	GA	0	0	1	0	1	
	CA-GA	0	0	0	0	1	
	PCA-GA	0	0	0	0	0	
بیشترین اختلاف با جواب	GA	0	1	1	4	13	
	CA-GA	0	0	1	3	10	
	PCA-GA	0	0	1	1	11	
مجموع اختلافات با جواب	GA	0	1	15	7	70	
	CA-GA	0	0	2	4	44	
	PCA-GA	0	0	2	1	43	

(د)

جدول (2) : نتایج بدست آمده از تعداد اجراهای متفاوت با تعداد

جمعیت های اولیه 16.64 و 100

7- نتیجه گیری

الگوریتم ژنتیک یکی از معروفترین الگوریتم های تکاملی است که در حل مسائل Np-hard کاربرد زیادی دارد در این مقاله از ترکیب اتوماتای سلولی احتمالی با الگوریتم ژنتیک جهت بهبود الگوریتم ژنتیک استفاده شده است و برای سنجش کارایی الگوریتم ترکیبی آنرا بر روی پنج مسئله آزمایش و با الگوریتم ژنتیک و ژنتیک سلولی مقایسه کردیم نتایج نشاندهنده موفقیت الگوریتم ارائه شده نسبت به الگوریتم های مقایسه شده بود.

مراجع

- [۱] Frankel, David S., Model Driven Architecture: Applying MDA to Enterprise Computing, OMG Press, Wiley Publishing, 2003.
- [۲] Sannella, M. J., Constraint Satisfaction and Debugging for Interactive User Interfaces, Ph.D. Thesis, University of Washington, Seattle, WA, 1994.
- [۳] Zachman, John A., "A Framework for Information Systems Architecture", IBM Systems Journal, Vol. 26, No. 3, 1987.
- [۴] Plamondon, R., and Lorette, G., "Automatic Signature Verification and Writer Identification - The State of the Art", Pattern Recognition, Vol. 22, pp. 107-131, 1989.
- [۵] Object Management Group. Unified Modeling Language: Superstructure, Version 2.0, ptc/03-07-06, July 2003, <http://www.omg.org/cgi-bin/doc?ptc/2003-08-02>.