

## تولید الگوهای نماینده به کمک الگوریتم AD-AIRS

امین زارع

دانشجوی کارشناسی ارشد هوش و مصنوعی و رباتیک

بخش علوم و مهندسی کامپیوتر، دانشگاه شیراز

aminzare@cse.shirazu.ac.ir

منصور ذوالقدری جهرمی

عضو هیئت علمی، بخش علوم و مهندسی کامپیوتر

دانشگاه شیراز

zjahromi@shirazu.ac.ir

جلوگیری می‌کند، همچنین با کاهش حجم داده‌ها، باعث افزایش سرعت و دقت طبقه‌بندی الگوریتم نزدیکترین همسایه می‌شود. AIRS بدلیل استفاده از KNN، به معیار فاصله حساس است. این حساسیت در مورد مجموعه‌های کوچک الگوهای نماینده محسوس‌تر است. در این روش طبقه‌بندی، برای بهبود معیار فاصله از روش‌های یادگیری فاصله استفاده می‌شود. این روش‌ها به کمک وزن‌دهی به نمونه‌ها سعی در بهبود معیار فاصله دارند. فرض کنید مجموعه‌ی  $T = \{x_1, x_2, \dots, x_n\}$  شامل الگوهای آموزشی با  $d$  ویژگی باشند و  $Q$  الگوی ورودی است. فاصله اقلیدسی بین دو الگوی  $x_i$  و  $Q$  از رابطه‌ی (۱) محاسبه می‌شود.

$$D(Q, x_i) = \sqrt{\sum_{k=1}^d (Q_k - x_{ki})^2} \quad (1)$$

هنگامی که از روش‌های یادگیری فاصله استفاده می‌شود، وزن  $w$  بسته به روش مورد استفاده، در رابطه (۱) وارد می‌گردد. انواع مختلفی از روش‌های وزن‌دهی وجود دارد [5,6].

گروهی از روش‌های وزن‌دهی، به هر الگو وزن خاصی نسبت می‌دهند. که در این حالت رابطه‌ی (۱) به صورت رابطه‌ی (۲) تغییر پیدا می‌کند. روش ارائه شده در این مقاله از رابطه‌ی (۲) استفاده می‌کند.

$$D(Q, x_i) = w_i \sqrt{\sum_{k=1}^d (Q_k - x_{ki})^2} \quad (2)$$

در ادامه در بخش دوم اصول سیستم ایمنی استفاده شده در AIRS توضیح داده می‌شود و الگوریتم طبقه‌بندی کننده‌ی AIRS در بخش سوم معرفی می‌گردد. در بخش‌های ۴ و ۵ تغییرات روی الگوریتم AIRS توضیح داده می‌شود و الگوریتم AD-AIRS ارائه می‌گردد. نتیجه‌ی اجرای الگوریتم AD-AIRS و مقایسه‌ی آن با AIRS به همراه مقدار کاهش حجم داده‌ها بر روی مجموعه داده‌های استاندارد در بخش ۶ آمده است.

**چک بیده:** الگوریتم AIRS (Artificial Immune Recognition System)

با استفاده از مجموعه داده‌های آموزشی و با الهام گرفتن از سیستم ایمنی بدن سعی در ساختن الگوهای نماینده (یا سلول‌های حافظه) دارد. در فاز عمومیت، به کمک الگوریتم K نزدیکترین همسایه (KNN) و با استفاده از الگوهای نماینده ساخته شده، طبقه‌بندی داده‌های ورودی جدید انجام می‌پذیرد. تحقیقات اخیر نشان داده است که کارایی این روش طبقه‌بندی تا حد زیادی به معیار فاصله مورد استفاده وابسته است؛ در این مقاله، نسخه‌ای از الگوریتم AIRS به نام Adaptive Distance AIRS (AD-AIRS) ارائه می‌شود که از یک نوع معیار فاصله وفقی استفاده می‌کند. الگوریتم AD-AIRS در مقایسه با الگوریتم AIRS نه تنها از دقت بهتری برخوردار است بلکه تعداد الگوهای نماینده ساخته شده توسط آن کمتر از الگوریتم AIRS می‌باشد این مسئله از این لحاظ حائز اهمیت است که باعث افزایش سرعت در فاز طبقه‌بندی می‌شود.

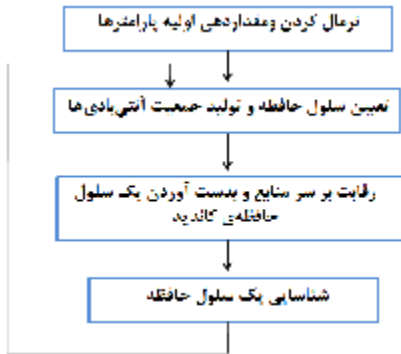
**واژه های کلیدی:** الگوهای نماینده، نزدیکترین همسایه، AIRS (Artificial Immune Recognition System)، فاصله‌ی وفقی، وزن‌دهی.

### ۱- مقدمه

در سال‌های اخیر، در زمینه‌های مختلف علوم کامپیوتر، تحقیقات وسیعی در مورد سیستم‌های ایمنی مصنوعی انجام گرفته است. با الهام گرفتن از توانایی‌های سیستم ایمنی بدن، الگوریتم‌هایی در زمینه‌ی شبکه‌های کامپیوتری [1]، خوشه‌بندی و طبقه‌بندی [2,3] و دیگر زمینه‌های علوم کامپیوتر [4] ارائه شده است.

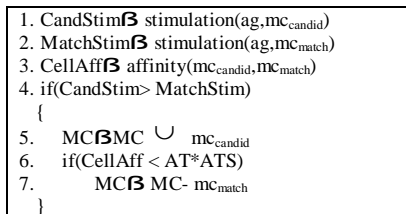
الگوریتم AIRS با توجه به سیستم ایمنی بدن، روشی برای یادگیری با ناظر ارائه کرده است. این الگوریتم با الهام گرفتن از سیستم ایمنی، الگوهای نماینده (یا همان سلول‌های حافظه) را تولید می‌کند و سپس با استفاده از طبقه‌بندی کننده‌ی KNN داده‌های تست را طبقه‌بندی می‌کند. در واقع الگوریتم AIRS با رفع مشکلات الگوریتم KNN دقت طبقه‌بندی را بهبود می‌بخشد. این الگوریتم به وسیله‌ی تولید الگوهای نماینده، مرز همواری بین کلاس‌ها ایجاد و از فرایادگیری

در مرحله ی سوم،  $mc_{match}$  تحت تکثیر و جهش قرار می گیرد و جمعیتی از آنتی بادی ها تولید می شود. این آنتی بادی ها تحت مکانیزمی به نام رقابت بر سر منابع، بهبود می یابند و نهایتاً آنتی بادی که نسبت به تحریک آنتی ژن ورودی حساس تر است به عنوان سلول حافظه ی کاندید ( $mc_{candid}$ ) انتخاب می شود.



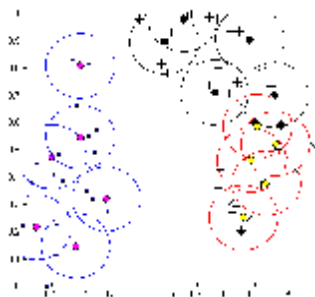
شکل ۱- مراحل مختلف الگوریتم AIRS

مرحله آخر الگوریتم آموزش، شامل شناسایی توانایی  $mc_{candid}$  در مقابل  $mc_{match}$  برای قرارگیری و یا جایگزینی در جمعیت سلول های حافظه است. اگر  $mc_{match}$  در شعاع همسایگی  $mc_{candid}$  قرار دارد آن را از مجموعه سلول های حافظه حذف می کنیم (خطوط ۶،۷ در شکل ۲). پارامتر  $ATS*AT$  برای هر سلول حافظه مانند یک شعاع همسایگی عمل می کند. آستانه شباهت  $AT$  و پارامتر  $ATS$  که توسط کاربر تعیین می شود ( $0 \leq ATS \leq 1$ ) اثر مهمی بر روی دقت طبقه بندی و تعداد سلول های حافظه ی بدست آمده دارد. شکل (۲) روال این مرحله را نشان می دهد.



شکل ۲- روال شناسایی سلول حافظه در الگوریتم AIRS

در شکل (۳)، الگوریتم AIRS بر روی یک مجموعه داده ی سه کلاسه با دو ویژگی اعمال و سلول های حافظه ی بدست آمده و شعاع همسایگی آن ها نشان داده شده است.



## ۲- اصول سیستم ایمنی بدن

بدن همه ی موجودات زنده، مکانیزم های خاصی برای دفاع از بدن، و تشخیص عوامل خارجی حمله کننده به آن را دارد و می تواند عوامل خارجی را که قبلاً با آنها مواجه شده است به خاطر آورد. به همین دلیل گاهی به آن مغز دوم نیز گفته می شود. عناصری که توسط سیستم ایمنی بدن تشخیص داده می شوند، «آنتی ژن» ( $Ag$ ) نام دارند. سلول های ایمنی بدن، کار دفاع در مقابل آنتی ژن های غیر خودی ورودی به بدن را به عهده دارند. سلول های ایمنی بدن به دو دسته ی سلول های  $T$  و  $B$  تقسیم می شوند. گیرنده های روی سطح سلول های  $T$  را  $TCR$  و گیرنده های روی سطح سلول های نوع  $B$  را «آنتی بادی» ( $Ab$ ) می نامند. الگوریتم AIRS بر اساس نحوه ی کار آنتی بادی ها عمل می کند.

## ۳- طبقه بندی کننده AIRS

AIRS اولین الگوریتم طبقه بندی الهام گرفته شده از سیستم ایمنی بدن می باشد، که در سال ۲۰۰۱ توسط Watkins ارائه شد [2]. از نگاه داده کاوی، AIRS الگوریتمی بر پایه خوشه بندی است که به کمک آن کار طبقه بندی را انجام می دهد. AIRS با یادگیری داده های ورودی، بردارهایی را به عنوان مراکز خوشه ها (الگوهای نماینده یا همان سلول های حافظه) پیدا می کند و با استفاده از طبقه بندی کننده  $K$  نزدیکترین همسایه کار طبقه بندی را انجام می دهد. نکته ی جالب در AIRS آن است که به عنوان یک الگوریتم طبقه بندی، سعی در پیدا کردن تعداد و مکان بهینه ی مراکز خوشه ها دارد.

در الگوریتم AIRS با سه جمعیت متفاوت سر و کار داریم. جمعیت آنتی ژن ها یا داده های آموزشی ( $ag_i$  ها)، جمعیت آنتی بادی ها ( $ab_i$  ها) و جمعیت سلول های حافظه ( $mc_i$  ها). آنتی بادی ها و سلول های حافظه بردارهایی از فضای ویژگی همراه با برجسب کلاس خود می باشند. جمعیت  $m$  عضوی از سلول های حافظه به صورت  $MC = \{mc_1, \dots, mc_m\}$  نشان داده می شوند.

## ۳-۱- الگوریتم AIRS

در این قسمت مراحل مختلف الگوریتم AIRS را که شرح کامل آن در [2] آمده است، به طور خلاصه توضیح می دهیم. الگوریتم AIRS شامل ۴ مرحله اصلی می باشد که در شکل (۱) نشان داده شده است. در مرحله ی اول داده ها نرمال می شوند و بعد از آن مقدار آستانه شباهت ( $AT$ ) (میانگین فاصله ی بین داده های آموزشی) بدست می آید. این مقدار در مرحله ی (۴) استفاده می شود. در مرحله ی دوم، بر اساس رابطه ی (۳)، سلول حافظه ای که دارای بیشترین تحریک نسبت به آنتی ژن ورودی در حال آموزش است پیدا می شود. به این سلول «سلول حافظه ی تطبیق شده» ( $mc_{match}$ ) می گوئیم.

$$Stimulation(ag, mc) = 1 - D(ag, mc) \quad (3)$$

می کند. شکل (۵) نتیجه اجرای AD-AIRS را روی مجموعه داده های شکل (۳) نشان می دهد.

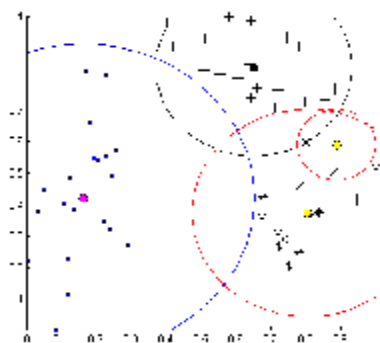
### ۵- بهبود الگوریتم AD-AIRS با میانگین گیری سلول های حافظه

می توان با استفاده از مفهوم خوشه بندی کارایی الگوریتم AD-AIRS را در پیدا کردن الگوهای نماینده و کاهش حجم داده ها بهبود بخشید. در الگوریتم AD-AIRS، هر کدام از سلول های حافظه به عنوان مراکز خوشه های درون کلاسی می باشند و اگر سلول حافظه ای در شعاع دیگری قرار گرفت، مرکز خوشه، توسط میانگین وزنی این دو سلول حافظه به روز می شود. (مانند الگوریتم خوشه بندی K میانگین) (شکل ۶)

```

1. Radius of mcmatch is calculated from previous iteration
2. Calculate radiuses of mccandid
3. CandStim $\beta$  weighted_stimulation(ag,mccandid)
4. MatchStim $\beta$  weighted_stimulation(ag,mcmatch)
5. CellAff $\beta$  affinity(mccandid,mcmatch)
6. if(CandStim > MatchStim)
7.   if( CellAff > mcmatch .r) // r is radiuses of mcmatch
8.     MCBMC  $\cup$  mccandid
    
```

شکل ۴- روال شناسایی سلول حافظه در الگوریتم AD-AIRS



شکل ۵- نتیجه اجرای الگوریتم AD-AIRS و سلول های حافظه ای بدست آمده با شعاع متغیر

```

1. Radius of mcmatch is calculated from previous iteration
2. Calculate radiuses of mccandid
3. CandStim $\beta$  weighted_stimulation(ag,mccandid)
4. MatchStim $\beta$  weighted_stimulation(ag,mcmatch)
5. CellAff $\beta$  affinity(mccandid,mcmatch)
6. if(CandStim > MatchStim)
7.   if( CellAff < mcmatch .r) // r is radiuses of mcmatch
8.     {
9.       mcaverage = (wmatch* mcmatch + wcandid*mccandid)/(wmatch + wcandid)
10.      Waverage = wmatch+1
11.      Calculate radiuses of mcaverage
12.      MCBMC  $\cup$  mcaverage
13.      MCB MC- mcmatch
14.     }
    
```

شکل ۶- روال شناسایی سلول حافظه در الگوریتم AD-AIRS به کمک میانگین گیری وزنی

شکل ۳- نتیجه اجرای الگوریتم AIRS و سلول های حافظه ای بدست آمده با شعاع همسایگی ثابت

### ۳-۲- طبقه بندی

بعد از پایان مرحله آموزش، مجموعه سلول های حافظه (الگوهای نماینده)، مجموعه ای کاهش یافته از داده های اصلی هستند. از این مجموعه برای طبقه بندی داده های تست، توسط الگوریتم K نزدیکترین همسایه استفاده می شود.

### ۴- الگوریتم AD-AIRS

در این مقاله، برای کم کردن حساسیت الگوریتم AIRS به معیار فاصله و بهبود دقت آن، در مرحله طبقه بندی از یک نوع معیار فاصله وقتی محلی استفاده شده است. این کار به کمک وزن دهی به الگوهای نماینده انجام می شود. همچنین در مرحله تعیین سلول حافظه، تابع تحریک تغییر می کند و نهایتاً از نوعی مکانیزم شعاع متغیر وقتی برای هر یک از الگوهای نماینده استفاده شده است و نسخه ای از AIRS به نام AD-AIRS ارائه می گردد. در الگوریتم AD-AIRS از مکانیزم وزن دهی که Wang در [7] معرفی کرده، استفاده شده است.

برای وزن دهی به الگوی نماینده  $mc$ ، بزرگترین دایره به مرکز  $mc$ ، که هیچ الگویی از کلاس مخالف را شامل نمی شود، در نظر گرفته و شعاع  $r$  محاسبه می شود و وزن  $w_{mc}$  با  $r$  نسبت عکس دارد. با داشتن وزن الگوی نماینده  $mc$  فاصله ای داده ای  $z$  با این الگو از رابطه (۴) بدست می آید:

$$D_{new}(mc, z) = D(mc, z) \times w_{mc} \quad (4)$$

استفاده از این فاصله وقتی باعث می شود یک داده ای تست به الگوهای نماینده مرکز کلاسها (با شعاع بزرگتر) نزدیکتر از الگوهای نماینده مرزی در نظر گرفته شود. با استفاده از این شیوهی وزن دهی در مرحله تعیین سلول حافظه، تابع تحریک به صورتی تغییر داده می شود که الگوهای نماینده ای که نزدیک مرزها واقع شده اند، با احتمال کمتری به مجموعه سلول های حافظه اضافه شوند. (رابطه ۵))

$$Weighted\_Stimulation(ag, mc) = \frac{1}{(D(ag, mc) * mc.weight)} \quad (5)$$

از دیگر اشکالات الگوریتم AIRS، استفاده از شعاع همسایگی ثابت، برای هر یک از الگوهای نماینده است. برای رفع آن، شعاع همسایگی هر الگوی نماینده، فاصله ای آن تا نزدیکترین الگوی دشمن (الگو از کلاس مخالف) در نظر گرفته می شود. چون هر الگوی نماینده، فضای بیشتری را پوشش می دهد با احتمال کمتری در این فضا نماینده دیگری ساخته می شود و به این ترتیب حجم داده ها نیز کمتر می شود. این تغییر سبب حذف پارامتر ATS از الگوریتم می شود. بدین ترتیب، مرحله شناسایی سلول حافظه به صورت روال شکل (۴) تغییر پیدا

جدول ۱- مقایسه الگوریتم AD-AIRS و AIRS و نزدیکترین همسایه

	Nearest Neighbor(k=1)		AIRS			AD-AIRS (k=1)	
	Accuracy	%	AIRS (k=1) Accuracy	AIRS(k=7) Accuracy	%	Accuracy	%
Iris	0.956	1	0.96	0.952	0.35	<b>0.970</b>	<b>0.199</b>
Glass	0.735	1	0.7	0.715	0.83	<b>0.74</b>	<b>0.213</b>
Diabetes	0.685	1	0.674	0.707	0.68	<b>0.738</b>	<b>0.127</b>
Bupa	0.599	1	0.619	0.638	1.34	<b>0.66</b>	<b>0.232</b>
Ionosphere	0.868	1	0.869	0.867	0.70	<b>0.951</b>	<b>0.330</b>
Sonar	0.827	1	0.841	0.824	0.93	<b>0.864</b>	<b>0.220</b>
WBC	0.955	1	0.961	0.962	0.66	<b>0.972</b>	<b>0.036</b>
Waveform	0.771	1	0.753	0.813	0.77	<b>0.822</b>	<b>0.087</b>
German	0.675	1	0.68	0.703	0.76	<b>0.728</b>	<b>0.14</b>
Average	0.785	1	0.784	0.797	0.78	<b>0.827</b>	<b>0.176</b>

نوعی شعاع وفقی استفاده می‌شود و همچنین استفاده از مکانیزم‌های تشویق و تنبیه و خوشه‌بندی باعث می‌شود الگوهای نماینده در مکان‌های بهتری پیدا شوند. تمام این تغییرات موجب بهبود نتایج نسبت به نسخه AIRS ساده و الگوریتم نزدیکترین همسایه شده است. ضمن اینکه کاهش حجم داده‌ها نیز قابل توجه می‌باشد.

#### ۸- مراجع

- [1] Swimmer, M., *Using the danger model of immune systems for distributed defense in modern data networks*, Computer Networks, Vol. 51, pp. 1315-1333, April 2007.
- [2] Watkins, A., *AIRS: A resource limited artificial immune classifier*, M.S. thesis, Mississippi State University, December, 2001.
- [3] Watkins, A., Timmis, J., *Artificial Immune Recognition System (AIRS): Revisions and Refinements*. ICARIS :1st International Conference on Artificial Immune Systems, pp. 173-181, September, 2002.
- [4] Schadwinkel, S., Dilger, W., *A Dynamic Approach to Artificial Immune Systems utilizing Neural Networks*. Proceedings of the 8th annual conference on Genetic and evolutionary computation (GECCO), pp. 131-132, Seattle, July, 2006.
- [5] Paredes, R., Vidal, E., *Learning prototypes and distances: A prototype reduction technique based on nearest neighbor error minimization*, Pattern Recognition Lett 39, 180 – 188, 2006
- [6] Kohavi, R., Langley, P., Yung, Y., *The utility of feature weighting in nearest-neighbor algorithms*, Proceedings of the Ninth European Conference of Machine Learning, Prague, Springer, Berlin, 1997.
- [7] Wang, J., Neskovic, P., Cooper, L. Improving nearest neighbor rule with a simple adaptive distance measure, Pattern Recognition Letters 28, 207–213, 2007
- [8] Bezdek, J.C., Kuncheva, L.I. *Nearest Prototype Classifier Designs: An Experimental Study*, Int'l J. Intelligent Systems, vol. 16, pp. 1445-1473, 2001.

در میانگین‌گیری وزنی، وزن نماینده‌ای که هر بار در فرایند میانگین‌گیری شرکت می‌کند افزایش می‌یابد. این کار موجب تاثیر بیشتر نماینده‌های مهم‌تر، در تعیین مکان الگوهای نماینده می‌شود. میانگین‌گیری وزنی و به روز رسانی مکان نماینده‌ها، در واقع همان مکانیزم تشویق و تنبیه می‌باشد که در بسیاری از روش‌های دیگر مانند LVM و DSM [8] نیز استفاده می‌شود.

#### ۶- نتایج

برای ارزیابی کارایی، الگوریتم AD-AIRS، بر روی مجموعه داده‌های UCI آزمایش شده است. برای پیدا کردن دقت از روش ده قسمتی (10-fold cross validation) استفاده شده است و دقت‌های ارائه شده با سه بار تکرار این عمل می‌باشد. در جدول (۱)، نتایج الگوریتم ارائه شده و الگوریتم AIRS به ازای  $k=1$  و  $k=7$  که در [9] نیز آمده است مقایسه‌ی شده‌اند. الگوریتم AD-AIRS به دلیل استفاده از فاصله‌ی وفقی، نیاز به تعیین  $k$  بهینه ندارد. نتایج ستون % بیان‌کننده‌ی حجمی از داده‌ها است که در طبقه‌بندی نهایی از آن استفاده می‌شود. در جدول (۱) نتایج بهتر به صورت پر رنگ در آمده است.

#### ۷- نتیجه‌گیری

در این مقاله الگوریتم AD-AIRS معرفی شده است. این الگوریتم بهبود یافته‌ی الگوریتم AIRS می‌باشد که از سیستم ایمنی بدن الهام گرفته است. الگوریتم AD-AIRS با تولید الگوهای نماینده (سلول‌های حافظه) و با استفاده از الگوریتم نزدیکترین همسایه، داده‌های تست را طبقه‌بندی می‌کند. برای کم کردن حساسیت الگوریتم AD-AIRS به معیار فاصله و بهبود دقت آن، از نوعی معیار فاصله وفقی محلی استفاده شده است. این کار به کمک وزن‌دهی به الگوهای نماینده انجام می‌شود. برای کاهش بیشتر حجم داده‌ها، برای هر یک از الگوهای نماینده از

- [9] Meng,L., Putten, P., Wang, H., A Comprehensive Benchmark of the Artificial Immune Recognition System (AIRS) , Lecture Notes in Computer Science ,Springer Berlin / Heidelberg , Volume 3584/2005 , 2005