مجله مدل‌سازی پیشرفته ریاضی
سال xxxx، دوره xx، شماره x، ص xxx–xxx

شاپا چاپی: ۲۲۵۱-۸۰۸۸
DOI: xxxx

مقایسه روش‌های تجزیهء مود دینامیکی دقیق و بهینه‌سازی تفکیک مؤلفه‌های خطی و غیرخطی در مدل‌سازی داده‌محور رشد تومور

سکینه اسمعیلی [(۱)] [۱] و نجمه منسوجیان [(۱)]

[(۱)] دانشکده علوم ریاضی، گروه ریاضی کاربردی، دانشگاه تربیت مدرس، تهران، ایران

دبیر مسئول:

تاریخ دریافت: xxx تاریخ پذیرش: xxx

چکیده: در این مقاله، عملکرد دو روش تحلیل دینامیکی داده‌محور $exDMD$ و $LANDO$، با استفاده از داده‌های به‌دست آمده از یک سیستم دینامیکی که مدلی از رشد تومور است، مقایسه شده‌اند. این داده‌های حاصل از مدل رشد تومور شامل ۳۳ ویژگی هستند. در این مجموعه‌داده، ویژگی‌ها بیانگر چگالی مقیاس‌شدهء سلول‌های تکثیرشونده در مکان‌های مختلف درون تومور و نمونه‌ها نمایانگر تغییرات زمانی چگالی سلولی در بازه‌های زمانی متفاوت می‌باشند. روش $exDMD$ با رویکردی خطی، در تحلیل سیستم‌های تقریباً خطی عملکرد مناسبی دارد، اما در مواجهه با رفتارهای غیرخطی دارای محدودیت است. در مقابل، $LANDO$ با بهره‌گیری از کرنل و دیکشنری‌های تنک، امکان شناسایی مؤثر اجزای خطی و غیرخطی سیستم را فراهم می‌کند. در این مطالعه، پس از تشریح مفهومی و ریاضی هر دو روش، نتایج حاصل از پیاده‌سازی آن‌ها بر روی داده‌های حاصل از مدل ریاضی رشد تومور در محیط متلب و با اعمال سطوح مختلف نویز به داده‌ها بررسی شده‌اند. مقایسهء خروجی‌ها نشان می‌دهد که روش $LANDO$ در مدل‌سازی رفتارهای غیرخطی دقت بالاتری دارد. حتی پیش‌بینی انجام شده با استفاده از عملگر خطی به‌دست‌آمده از طریق روش $LANDO$ نیز دقیق‌تر از $exDMD$ است. این نتایج می‌توانند به بهبود تحلیل و پیش‌بینی پویایی بیماری‌هایی مانند سرطان کمک کنند.

واژه‌های کلیدی: تجزیهء مود دینامیکی دقیق($exDMD$)، تجزیهء مقدار تکین، کرنل، بهینه‌سازی تفکیک مؤلفه‌های خطی و غیرخطی($LANDO$).

[۱] نویسنده مسئول مقاله

(Sakine Esmaili) sakine.esmaili@modares.ac.ir

۱ مقدمه

تجزیهء مود دینامیکی دقیق[۲] یک روش در تحلیل داده‌های زمانی است که به‌طور خاص برای استخراج دینامیک‌های زیربنایی[۳] در سیستم‌های پیچیده، توسعه یافته است. این روش برای نخستین‌بار [۱۰] در تحلیل جریان‌های سیال مطرح شد و از آن زمان تاکنون در حوزه‌های مختلفی به کار گرفته شده است.

ابتدا الگوریتم تجزیهء مود دینامیکی استاندارد (DMD) در سال ۲۰۰۸ توسط P. J. Schmid و J. Sesterhenn در مقالهء [۷] برای پیش‌بینی دینامیک سیستم‌ها معرفی شد. این الگوریتم مجموعه ترتیبی از بردارهای داده‌ای $\{z_0, \dots, z_m\}$، به گونه‌ای که هر $z_k \in \mathbb{R}^n$، را در نظر می‌گیرد که در آن ترتیب بردارهای z_k اهمیت دارد.

در مطالعهء انجام‌شده توسط Tu و همکاران در سال ۲۰۱۳، مبانی ریاضی روش تجزیهء مود دینامیکی توسعه داده شد و ارتباط این روش با سایر تکنیک‌های کاهش ابعاد، از جمله تجزیهء مقدار تکین[۴] ، به‌صورت نظام‌مند تبیین شده‌است [۱۰]. در الگوریتم تجزیهء مود دینامیکی دقیق، این محدودیت ترتیب زمانی، بر روی داده‌ها تسهیل شده و جفت‌های داده‌ای[۵] $\{(x_1, y_1), \dots, (x_m, y_m)\}$ در نظر گرفته شده‌است. سپس $exDMD$ بر اساس ماتریس‌های داده $n \times m$ تعریف می‌شود:

$$\mathbf{X} = \begin{bmatrix} x_1, & \cdots & , x_m \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} y_1 & \cdots & , y_m \end{bmatrix}.$$

روش $exDMD$ ابزاری مبتنی بر داده است که تلاش می‌کند بهترین مدل خطی با رتبه پایین را تقریب بزند. این روش یک ماتریس رتبه پایین بهینه به نام $\mathbf{A}$ را تعیین می‌کند به طوری که: $y_j \approx \mathbf{A}x_j$ برای $j = 1, \cdots, m$. عبارت فوق به‌صورت زیر بیان می‌شود:

$$\arg \min_{rank(\mathbf{A}) \leq r} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F.$$

یکی از کاربردهای برجستهء $exDMD$ در سال‌های اخیر، تحلیل داده‌های زیستی به‌ویژه داده‌های مرتبط با سرطان بوده است [۱۱]. با استفاده از $exDMD$، می‌توان الگوهای زمانی سلول‌های سرطانی را استخراج کرد که در تشخیص زودهنگام و پیش‌بینی رفتار تومورها کاربرد دارد. این روش، بدون نیاز به مدل‌سازی صریح، قادر است ساختارهای تکرارشونده و مؤلفه‌های اصلی دینامیکی در داده‌ها را شناسایی کند. بنابراین، $exDMD$ به عنوان پلی میان ریاضیات کاربردی و علوم زیستی، افق‌های جدیدی در تحلیل دینامیکی فرآیندهای زیستی گشوده است.

علیرغم کارایی بالای روش $exDMD$ در استخراج دینامیک‌های غالب از داده‌های زمانی، این روش با محدودیت‌هایی مواجه است. از جمله مهم‌ترین این محدودیت‌ها می‌توان به فرض خطی بودن دینامیک‌ها و حساسیت به نویز داده‌های ورودی اشاره کرد. همچنین، $exDMD$ در مدل‌سازی سامانه‌های غیرخطی پیچیده دچار افت دقت می‌شود و این فضای ویژگی خطی ممکن است قادر به نمایش مؤثر ساختارهای غیرخطی نهفته در داده نباشد.

یکی از راهکارهای مؤثر برای رفع این محدودیت‌ها، بهره‌گیری از توابع کرنل است که امکان نگاشت داده‌ها به فضای ویژگی با ابعاد بالاتر را فراهم می‌سازد و در نتیجه توانایی تحلیل دینامیک‌های غیرخطی را بهبود می‌بخشد.

این راهکار در الگوریتم بهینه‌سازی تفکیک مؤلفه‌های خطی و غیر خطی[۶] بکار گرفته شده است. این الگوریتم توسط $Baddoo$ و همکاران در سال ۲۰۲۲ ارائه شد [۱]. در مقالهء [۱]، با استفاده از روش $LANDO$ به تحلیل و مطالعه سیستم‌های دینامیکی شامل سیستم لورنز و معادله دیفرانسیل جزئی غیرخطی برگرز پرداخته شده‌است.

استفاده از کرنل‌های مناسب امکان استخراج الگوهای دینامیکی پنهان در فرآیندهای مولکولی، تکثیر سلولی و پاسخ به درمان را فراهم می‌سازد. مقالات [۵، ۹] نمونه هایی از کاربرد کرنل در حوزه بیماری سرطان هستند.

این مثال‌ها نشان می‌دهند که $LANDO$ ابزاری قدرتمند برای مدل‌سازی داده‌محور سیستم‌های دینامیکی پیچیده و غیرخطی است.

۲ الگوریتم‌های مدل‌سازی داده‌محور: exDMD و LANDO

در این بخش به معرفی دو الگوریتم داده محور که در این مقاله از آنها استفاده شده است، می پردازیم:

۱. الگوریتم تجزیه مود دینامیکی دقیق [۱۰]

---

[۲] Exact Dynamic Mode Decomposition - exDMD

[۳] Underlying Dynamics

[۴] Singular Value Decomposition

[۵] Data Pair دو مقدار اندازه گیری شده که با هم مرتبط هستند.

[۶] Linear and Nonlinear Disambiguation Optimization - LANDO

(آ) جفت های داده‌ای $\{(x_۱, y_۱), \ldots, (x_m, y_m)\}$ را در ماتریسهای X و Y به‌صورت زیر مرتب کنید:

$$\mathbf{X} = \begin{bmatrix} x_۱, & \cdots & , x_m \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} y_۱ & \cdots & , y_m \end{bmatrix}.$$

برای مجموعه داده‌های X و Y، عملگر $\mathbf{A}$ را به‌صورت زیر تعریف کنید:

$$\mathbf{A} \triangleq YX^{\dagger}, \tag{۱.۲}$$

که $X^{\dagger}$ ماتریس شبه معکوس X است.

(ب) تجزیهء مقدار تکین ماتریس X را محاسبه کنید:

$$X = U\Sigma V^{*},$$

که در آن U ماتریسی با ابعاد $n \times r$ و Σ ماتریسی با ابعاد $r \times r$ و V ماتریسی با ابعاد $m \times r$ است و رتبهء ماتریس X، r است.

(ج) ماتریس $\tilde{\mathbf{A}}$ را با استفاده از رابطهء زیر تعریف کنید:

$$\tilde{\mathbf{A}} \triangleq U^{*}YV\Sigma^{-۱}. \tag{۲.۲}$$

(د) مقادیر ویژه و بردارهای ویژهء ماتریس $\tilde{\mathbf{A}}$ با رابطهء زیر نشان داده می‌شوند:

$$\tilde{\mathbf{A}}\omega = \lambda\omega.$$

(ه) سپس مود $exDMD$ مربوط به مقدار ویژهء λ توسط رابطهء زیر محاسبه می‌شود:

$$\varphi \triangleq \frac{۱}{\lambda} YV\Sigma^{-۱}\omega.$$

قضیه **۱.۲**. [۱۰] هر جفت (φ, λ) که توسط الگوریتم تجزیهء مود دینامیکی دقیق تولید می‌شود، یک جفت (مقدار ویژه،بردار ویژه) از ماتریس $\mathbf{A}$ در رابطه (۱.۲) است. علاوه بر این، این الگوریتم تمام مقادیر ویژهء غیرصفر ماتریس $\mathbf{A}$ را شناسایی می‌کند.

اثبات. [۱۰] از تجزیهء مقدار تکین ماتریس X، می توانیم ماتریس شبه معکوس X را به‌صورت زیر بنویسیم:

$$X^{\dagger} = V\Sigma^{-۱}U^{*}.$$

بنابراین

$$\mathbf{A} = YV\Sigma^{-۱}U^{*} = BU^{*},$$

که در آن

$$B \triangleq YV\Sigma^{-۱}.$$

علاوه بر این می توانیم (۲.۲) را به‌صورت زیر بنویسیم:

$$\tilde{\mathbf{A}} = U^{*}YV\Sigma^{-۱} = U^{*}B.$$

حال فرض کنید $\tilde{\mathbf{A}}\omega = \lambda\omega$ با $\lambda \neq ۰$ و $\varphi = \frac{۱}{\lambda}B\omega$، بنابراین رابطه زیر برقرار است:

$$\begin{aligned} \mathbf{A}\varphi &= \frac{۱}{\lambda} BU^{*}B\omega \\ &= B\frac{۱}{\lambda}\tilde{\mathbf{A}}\omega \\ &= B\omega \\ &= \lambda\varphi. \end{aligned}$$

برای نشان‌دادن اینکه الگوریتم تجزیهء مود دینامیکی دقیق همهء مقادیر ویژهء غیر صفر $\mathbf{A}$ را شناسایی می‌کند، فرض کنید:

$$\mathbf{A}\varphi = \lambda\varphi, (\lambda \neq ۰).$$

بنابراین رابطهء زیر برقرار است:

$$U^*\mathbf{A}\varphi = \lambda U^*\varphi \rightarrow \tilde{\mathbf{A}}U^*\varphi = \lambda U^*\varphi$$

با در نظرگرفتن $U^*\varphi = \omega$ داریم:

$$\tilde{\mathbf{A}}\omega = \lambda\omega.$$

و این یعنی هر مقدار ویژهء غیر صفر $\mathbf{A}$ به یک مقدار ویژهء غیر صفر $\tilde{\mathbf{A}}$ مرتبط است. □

۲. الگوریتم بهینه‌سازی تفکیک مؤلفه‌های خطی و غیرخطی

در این بخش به ارائهء روش $LANDO$ می‌پردازیم که از کرنل‌ها برای ساخت مدل‌های داده‌محور استفاده می‌کند. هدف اصلی $LANDO$ جداسازی و شناسایی دقیق اجزای خطی و غیرخطی در دینامیک سیستم‌ها است [۱].

این روش با استفاده از رگرسیون کرنلی بر روی یک دیکشنری تنک[۷] از نمونه‌ها که به‌طور قابل توجهی در دینامیک سیستم مؤثر هستند، مدل‌سازی را انجام می‌دهد.

مسئله شامل یافتن تابعی به نام f است که داده‌های آموزشی را با رعایت معیارهایی از جمله قابلیت تعمیم[۸] و قابلیت تفسیر[۹] به‌درستی نگاشت کند، $f(x) \approx F(x)$. تابع F همان دینامیک واقعی سیستم است و ما داده‌هایی داریم که از آن حاصل شده‌اند. بنابراین، مدل تقریبی f را می‌سازیم که بسیاری از ویژگی‌های F را حفظ کند.

تابع f معمولاً به یک کلاس خاص از مدل‌ها محدود می‌شود. این مدل‌ها می‌توانند به‌صورت خطی، غیرخطی و چندجمله‌ای باشند. در ادامه مدل $f(x)$ را به‌صورت ترکیب خطی زیر نمایش می‌دهیم:

$$f(x) = \sum_{j=۱}^{N} \xi_j \phi_j(x) \quad \Rightarrow \quad f(x) = \boldsymbol{\Xi}\boldsymbol{\Phi}(\mathbf{x}).$$

در اینجا $\boldsymbol{\Phi}(\mathbf{x})$ شامل N مولفه برای توصیف دینامیک سیستم است که می‌توانند خطی یا غیرخطی باشند، مانند $\{۱, x, y, x^۲, xy, \sin(x), \cos(y), \dots\}$. $\boldsymbol{\Xi}$ ماتریسی از ضرایب است که مشخص می‌کند کدام‌یک از این عبارات در مدل فعال باشند و هر کدام با چه نسبتی (مقداری) نقش داشته باشند، در واقع کدام ترم‌های موجود در $\boldsymbol{\Phi}(\mathbf{x})$ واقعاً لازم‌اند. اگر یک ضریب صفر شود یعنی آن ترم در معادلهء نهایی حضور ندارد و اگر غیرصفر باشد یعنی آن ترم در مدل فعال است و ضریب، مقدار اهمیتش را نشان می‌دهد [۱].

هدف این است که $\boldsymbol{\Xi}$ را طوری پیدا کنیم که خطای مدل کمینه شود:

$$\arg\min_{\boldsymbol{\Xi}} \|Y - \boldsymbol{\Xi}\boldsymbol{\Phi}(\mathbf{x})\|_F + \lambda R(\boldsymbol{\Xi}).$$

در اینجا $\|Y - \boldsymbol{\Xi}\boldsymbol{\Phi}(\mathbf{x})\|_F$، نشان‌دهندهء خطای بین پیش‌بینی مدل و داده‌های واقعی است. همچنین $\lambda R(\boldsymbol{\Xi})$، یک منظم‌ساز است که باعث می‌شود مدل، بیش‌برازش نکند.

در مدل‌های پیچیده، تعداد ویژگی‌های N می‌تواند بسیار زیاد باشد، به‌طوری که حل مسئلهء بهینه‌سازی محاسباتی با مشکل روبرو شود. برای حل این مشکل، می‌توان به‌جای ویژگی‌های صریح $\boldsymbol{\Phi}(\mathbf{x})$ از یک تابع کرنل K، استفاده کرد:

$$f(x) = \sum_{j=۱}^{N} \xi_j \phi_j(x) = \sum_{i=۱}^{m} w_i K(x_i, x). \tag{۳.۲}$$

می‌توانیم رابطهء بالا را به‌صورت $f(x) = \mathbf{W}K(X, x)$ بنویسیم، بنابراین مسئلهء بهینه‌سازی به‌صورت زیر است:

$$\arg\min_{\mathbf{W}} \|Y - \mathbf{W}K(X, X)\|_F + \lambda R(\mathbf{W}). \tag{۴.۲}$$

---

[۷] Sparse Dictionary
[۸] Generalizability
[۹] Interpretability

از نظر تئوری، یک جواب برای (۴.۲) در صورت عدم وجود منظم‌سازی توسط شبه‌معکوس مور–پنفروز [۱۰] ارائه می‌شود [۱]:

$$\mathbf{W} = YK(X,X)^{\dagger}. \tag{۵.۲}$$

در مدل $LANDO$، این امکان وجود دارد که مؤلفهٔ خطی Lدینامیک‌ها به طور کارآمد استخراج شود. رابطه‌ای که برای این کار استفاده می‌شود به صورت زیر است [۱] :

$$Y = \mathbf{L}X + \mathbf{N}. \tag{۶.۲}$$

دیکشنری‌های تنک به ما کمک می‌کنند که به‌جای نگه‌داشتن تمام نمونه‌های آموزشی، تعداد کمی نمونهٔ مهم را انتخاب کنیم و مدل را فقط بر اساس آن‌ها بسازیم [۱]. ایدهٔ کلیدی این روش این است که مدل f تعریف‌شده در (۳.۲) را می‌توان به‌صورت تقریبی، برای یک انتخاب مناسب از $\tilde{X}$ که به‌عنوان دیکشنری شناخته می‌شود، به‌صورت زیر بیان کرد:

$$f(x) \approx \tilde{\mathbf{W}}K(\tilde{X},x), \tag{۷.۲}$$

در اینجا، نماد $\sim$ نشان می‌دهد که یک کمیت به دیکشنری مرتبط است.

سپس، مسئلهٔ بهینه‌سازی (۴.۲) را می‌توان به‌صورت تقریبی بیان کرد به‌طوری که

$$\arg\min_{\tilde{\mathbf{W}}} \|Y - \tilde{\mathbf{W}}K(\tilde{X},X)\|_F + \lambda R(\tilde{\mathbf{W}}).$$

یکی از روشهای افزایش قابلیت تفسیر مدل، تحلیل مؤلفهٔ خطی تابع f نسبت به یک حالت مرجع مشخص است. فرض می‌کنیم یک حالت پایهٔ [۱۱] $\bar{x}$، که می‌تواند میانگین داده‌ها، یک نقطه تعادلی یا صرفاً بردار صفر باشد در نظر گرفته شود. با تعریف $x' = x - \bar{x}$ ، تابع f به فرم زیر بازنویسی می‌شود [۱]:

$$f(x) = c + \mathbf{L}x' + \mathbf{N}(x'),$$

که در اینجا L یک اپراتور خطی، c یک مقدار ثابت و N یک اپراتور غیرخطی است که در رابطهٔ زیر صدق می‌کند:

$$\lim_{\|x'\|_۲ \to ۰} \frac{\|\mathbf{N}(x')\|_۲}{\|x'\|_۲} = ۰.$$

با در نظر گرفتن بسط تیلور f، رابطهٔ زیر برقرار است:

$$c = f(\bar{x}), \qquad \mathbf{L} = \nabla f(x)\big|_{x=\bar{x}}$$

و تابع باقی‌مانده (یا خطای تقریب) به شکل زیر تعریف می‌شود:

$$\mathbf{N}(x') = f(x) - f(\bar{x}) - \nabla f(x)\big|_{x=\bar{x}} x'.$$

با توجه به رابطهٔ (۷.۲):

$$\mathbf{L} = \nabla f(x)\big|_{x=\bar{x}} = \tilde{\mathbf{W}}\nabla K(\tilde{X},\bar{x}).$$

در بخش نتایج عددی نشان داده شده است که $LANDO$ عملکرد بهتری در مقایسه با $exDMD$ دارد. در ادامه، علاوه بر $\tilde{\mathbf{W}}K(\tilde{X},x)$، از قسمت خطی آن نیز طبق رابطهٔ زیر برای پیش‌بینی استفاده شده است:

$$f(\bar{x}) + \nabla f(x)\big|_{x=\bar{x}}(x - \bar{x}),$$

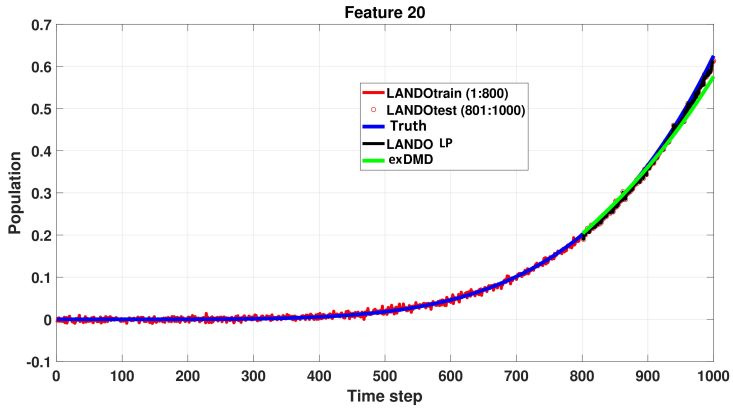
که در اینجا پیش‌بینی بر اساس قسمت خطی را با $LANDO\ LP$ نمایش می‌دهیم.

[۱۰] Moore Penrose Pseudoinverse
[۱۱] Base State

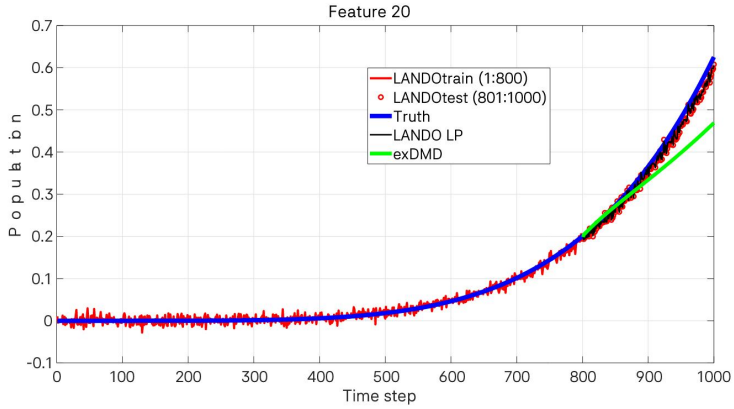
۳ نتایج عددی

مثال **۱.۲**. برای مقایسه عملکرد پیش‌بینی دو روش $exDMD$ و $LANDO$، فرآیند آموزش و ارزیابی مدل‌ها با استفاده از نرم‌افزار $MATLAB$ انجام شده و نتایج حاصل در شکل‌های زیر ارائه شده‌اند. مجموعه‌ی داده‌ها شامل ۱۰۰۰ نمونه با ۳۳ ویژگی برگرفته از یک مدل ریاضی رشد تومور است [۳، ۴]. در این مجموعه داده، ویژگی‌ها، چگالی مقیاس‌شدهء سلول‌های تکثیرشونده[۱۲] در مکان‌های مختلف درون تومور هستند و تعداد نمونه‌ها بیانگر چگالی سلول‌ها در زمان‌های مختلف است. در این آزمایش، تعداد ۸۰۰ نمونه ابتدایی به عنوان داده‌های آموزشی و ۲۰۰ نمونهء پایانی جهت ارزیابی دقت پیش‌بینی به‌کار گرفته شده‌اند.

در شکل‌های ۱، ۲ و ۳ برای نویزهای مختلف، عملکرد $exDMD$ و $LANDO$ و همچنین $LANDO\ LP$ نمایش داده شده‌اند. همان‌طور که در نمودارها مشاهده می‌شود، روش $LANDO$ توانسته است بازسازی دقیق‌تری از رفتار واقعی سیستم ارائه دهد. اختلاف عملکرد سه روش در داده‌های آزمون، گویای توانایی بالاتر روش $LANDO$ در مدل‌سازی است.

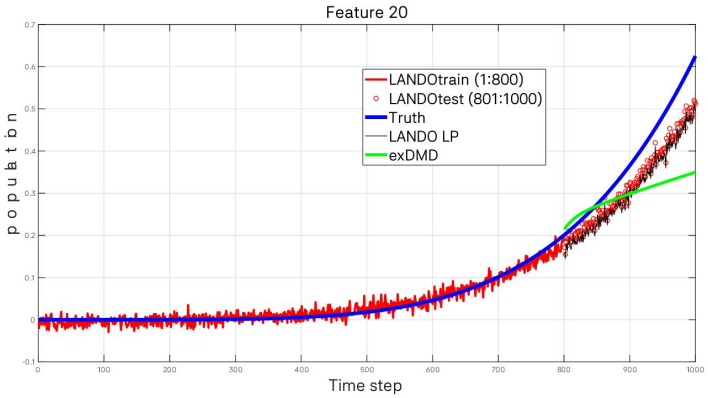


شکل ۱: مقایسه عملکرد پیش‌بینی مدل‌های LANDO، exDMD و LANDO LP با استفاده از داده‌های دارای نویز گوسی با انحراف معیار 0.01.



شکل ۲: مقایسه عملکرد پیش‌بینی مدل‌های LANDO، exDMD و LANDO LP با استفاده از داده‌های دارای نویز گوسی با انحراف معیار 0.02.

[۱۲] Proliferative Cells



شکل ۳: مقایسه عملکرد پیش‌بینی مدل‌های LANDO، exDMD و LANDO LP با استفاده از داده‌های دارای نویز گوسی با انحراف معیار 0.03.

در ادامه با استفاده از دو معیار خطا، میانگین قدر مطلق خطا[۱۳] (MAE) و ریشهء میانگین مربعات خطا[۱۴] $(RMSE)$، به سنجش دقت مدل $LANDO$ برای پیش‌بینی داده‌های مختلف در حضور نویز می‌پردازیم، که این معیارهای خطا در مهندسی، علوم داده و یادگیری ماشین برای اندازه‌گیری میزان انحراف خروجی مدل از داده‌های واقعی مورد استفاده قرار می‌گیرند و به‌صورت زیر تعریف می‌شوند:

۱. فرمول محاسبه خطای MAE:

$$\mathrm{MAE} = \frac{۱}{n}\sum_{i=۱}^{n} |y_i - \hat{y}_i| \,.$$

۲. فرمول محاسبه خطای $RMSE$:

$$\mathrm{RMSE} = \sqrt{\frac{۱}{n}\sum_{i=۱}^{n} (y_i - \hat{y}_i)^۲}.$$

بر اساس نتایج دو جدول به‌دست‌آمده از ارزیابی خطای پیش‌بینی بر روی داده‌های رشد تومور، مشخص شد که مدل $LANDO$ عملکرد بهتری نسبت مدل $exDMD$ دارد. مقادیر معیارهای MAE و $RMSE$ برای $LANDO$ به‌طور محسوسی کمتر از این مقادیر برای سایر مدل‌ها بودند. این موضوع بیانگر توانایی بالای این مدل در یادگیری دقیق دینامیک‌های پنهان و پیچیده‌ی سیستم است. این ویژگی، $LANDO$ را به گزینه‌ای بسیار مناسب برای کاربردهای واقعی تبدیل می‌کند. همچنین روش $LANDO\ LP$ نسبت به $exDMD$ عملکرد بهتری داشته و خطای کمتری تولید کرده‌است.

Models	MAE	RMSE
LANDO Prediction	۷.۸۴۷e-۳	۹.۲۵۹e-۳
exDMD Prediction	۱.۲۹۸e-۲	۱.۹۰۱e-۲
LANDO LP Prediction	۸.۷۵e-۳	۱.۰۳۳e-۲

جدول ۱: مقایسه‌ی معیارهای خطای MAE و RMSE برای مدل‌های LANDO، exDMD و LANDO LP با استفاده از داده‌های دارای نویز گوسی با انحراف معیار 0.01.

[۱۳] Mean Absolute Error
[۱۴] Root Mean Squared Error

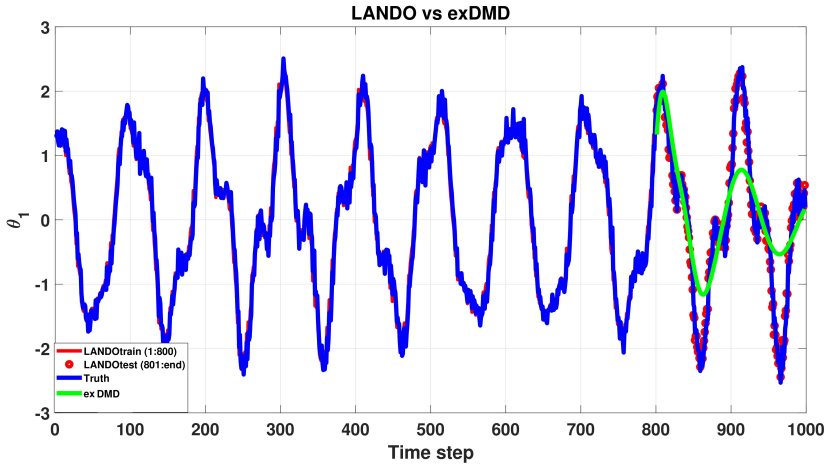
Models	MAE	RMSE
LANDO Prediction	۳.۰۶۲e-۲	۳.۴۵۷e-۲
exDMD Prediction	۴.۸۶۸e-۲	۶.۸۱۸e-۲
LANDO LP Prediction	۳.۳۵۵e-۲	۳.۶۹۴e-۲

جدول ۲: مقایسه‌ی معیارهای خطای MAE و RMSE برای مدل‌های LANDO، exDMD و LANDO LP با استفاده از داده‌های دارای نویز گوسی با انحراف معیار 0.02.

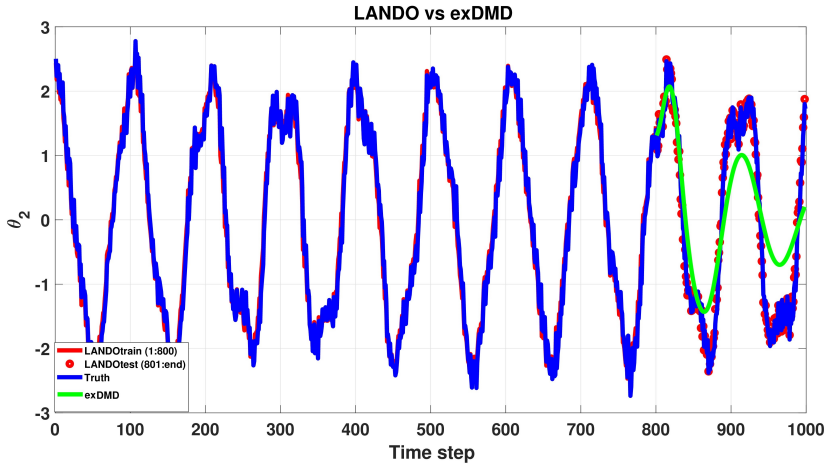
Models	MAE	RMSE
LANDO Prediction	۵.۹۶۱e-۲	۶.۶۲۷e-۲
exDMD Prediction	۹.۲۹۱e-۲	۱.۲۴۴e-۱
LANDO LP Prediction	۷.۱۹۹e-۲	۷.۷۶۸e-۲

جدول ۳: مقایسه‌ی معیارهای خطای MAE و RMSE برای مدل‌های LANDO، exDMD و LANDO LP با استفاده از داده‌های دارای نویز گوسی با انحراف معیار 0.03.

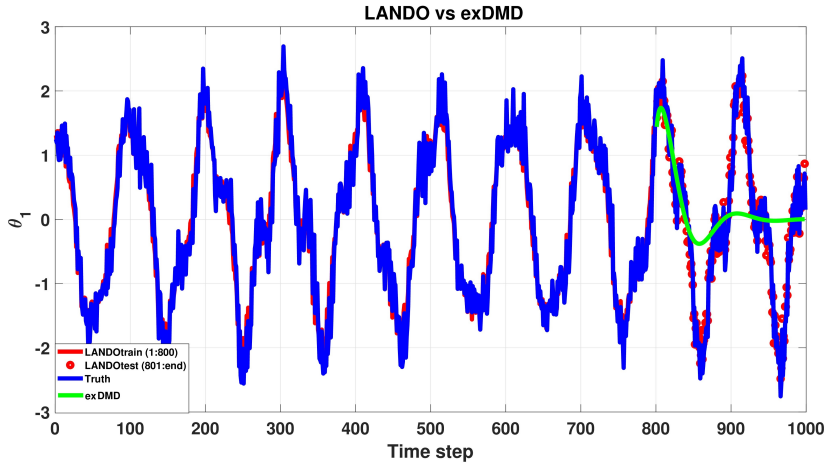
مثال **۲.۲**. در این مثال داده‌های مربوط به یک آونگ دوتایی با اضافه کردن نویز برای ارزیابی روش‌های $LANDO$ و $exDMD$ مورد استفاده قرار گرفته‌اند [۲، ۶، ۸]. برای مقایسه روش‌های $LANDO$ و $exDMD$، داده‌ها با نویزهای گوسی با انحراف معیار ۰٫۰۳ و ۰٫۰۶ آلوده شده‌اند. در شکل‌های ۴، ۵، ۶ و ۷ نتایج روش‌های $LANDO$ و $exDMD$ با داده‌ی اصلی مقایسه شده‌اند که در همه‌ی شکل‌ها نتایج $LANDO$ به داده‌ی اصلی نزدیکتر است.



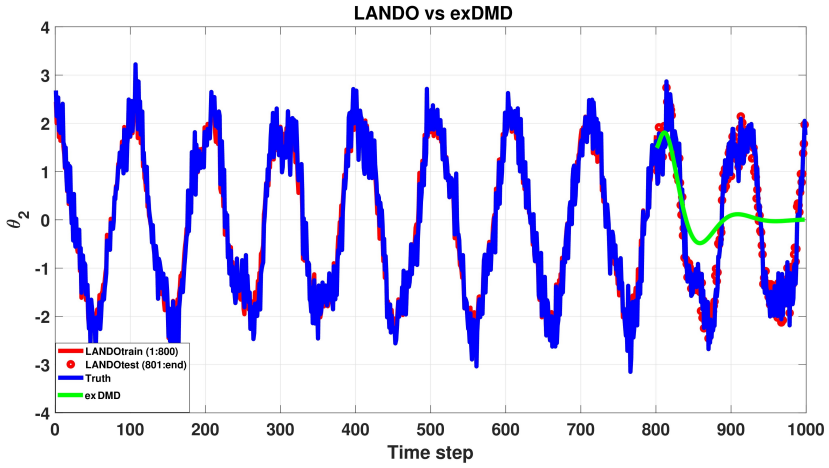
شکل ۴: مقایسه نتایج روش‌های LANDO و exDMD با داده‌های اصلی با استفاده از داده‌های آموزشی آلوده به نویز گوسی با انحراف معیار 0.03.



شکل ۵: مقایسه نتایج روش‌های LANDO و exDMD با داده‌های اصلی با استفاده از داده‌های آموزشی آلوده به نویز گوسی با انحراف معیار 0.03.



شکل ۶: مقایسه نتایج روش‌های LANDO و exDMD با داده‌های اصلی با استفاده از داده‌های آموزشی آلوده به نویز گوسی با انحراف معیار 0.06.



شکل ۷: مقایسه نتایج روش‌های LANDO و exDMD با داده‌های اصلی با استفاده از داده‌های آموزشی آلوده به نویز گوسی با انحراف معیار 0.06.

Models	MAE	RMSE
LANDO Prediction for $\theta_۱$	۹.۵۶۳e-۲	۱.۱۴۶e-۱
exDMD Prediction for $\theta_۱$	۵.۳۱۸e-۱	۷.۱۴۸e-۱
LANDO Prediction for $\theta_۲$	۱.۰۱۲e-۱	۱.۲۶۱e-۱
exDMD Prediction for $\theta_۲$	۶.۱۷۸e-۱	۷.۳۵۵e-۱

جدول ۴: مقایسه‌ی معیارهای خطای MAE و RMSE برای مدل‌های LANDO و exDMD با استفاده از داده‌های آموزشی آلوده به نویز گوسی با انحراف معیار 0.03.

Models	MAE	RMSE
LANDO Prediction for $\theta_۱$	۱.۳۵۵e-۱	۱.۶۲e-۱
exDMD Prediction for $\theta_۱$	۷.۶۹۵e-۱	۱.۰۵۶e۰
LANDO Prediction for $\theta_۲$	۱.۶۶۴e-۱	۱.۹۸۶e-۱
exDMD Prediction for $\theta_۲$	۱.۰۹۲e۰	۱.۲۳۳e۰

جدول ۵: مقایسه‌ی معیارهای خطای MAE و RMSE برای مدل‌های LANDO و exDMD با استفاده از داده‌های آموزشی آلوده به نویز گوسی با انحراف معیار 0.06.

۴ نتیجه‌گیری

نتایج به‌دست آمده از پیاده‌سازی دو روش $exDMD$ و $LANDO$ بر روی داده‌های حاصل از یک مدل رشد تومور، تفاوت‌های قابل‌توجهی را در دقت پیش‌بینی حالت سیستم نشان می‌دهند. روش $LANDO$ با بهره‌گیری از چارچوب کرنل و ساختار دیکشنری تنک، به‌طور موثرتری دینامیک‌های غیرخطی را شناسایی و بازسازی نمود. این نتایج نشان می‌دهند که $LANDO$ به‌ویژه در مدل‌سازی دینامیک‌های پیچیده زیستی، عملکرد قابل‌اتکاتری دارد. برای ارزیابی جامع‌تر، مثال آونگ دوتایی نیز مورد بررسی قرار گرفت. نتایج نشان دادند که $LANDO$ حتی در شرایط وجود نویزهای متفاوت نیز قابلیت بازسازی دینامیک سیستم را با دقت بالاتر حفظ کرده و نسبت به $exDMD$ عملکرد بهتری ارائه می‌دهد. به طور کلی، یافته‌های این پژوهش نشان می‌دهد که استفاده از رویکردهای مبتنی بر کرنل، $LANDO$، می‌تواند چشم‌انداز جدیدی در تحلیل و مدل‌سازی سیستم‌های غیرخطی و زیستی فراهم سازد و جایگزین مناسبی برای روش‌های کلاسیک همچون $exDMD$ باشد.

فهرست منابع

[1] Baddoo, Peter J, Herrmann, Benjamin, McKeon, Beverley J, and Brunton, Steven L. Kernel learning for robust dynamic mode decomposition: linear and nonlinear disambiguation optimization. *Proceedings of the Royal Society A*, 478(2260):20210830, 2022. doi:10.1098/rspa.2021.0830.

[2] Baddoo, Peter J, Herrmann, Benjamin, McKeon, Beverley J, Nathan Kutz, J, and Brunton, Steven L. Physics-informed dynamic mode decomposition. *Proceedings of the Royal Society A*, 479(2271):20220576, 2023. doi:10.48550/arXiv.2112.04307.

[3] Esmaili, Sakine and Eslahchi, MR. Application of collocation method for solving a parabolic-hyperbolic free boundary problem which models the growth of tumor with drug application. *Mathematical Methods in the Applied Sciences*, 40(5):1711–1733, 2017. doi:10.1002/mma.4092.

[4] Esmaili, Sakine and Eslahchi, MR. Numerical solution of optimal control problem for a model of tumour growth with drug application. *International Journal of Control*, 92(11):2712–2736, 2019. doi:10.1080/00207179.2018.1458159.

[5] Feng, Jie, Jiang, Limin, Li, Shuhao, Tang, Jijun, and Wen, Lan. Multi-omics data fusion via a joint kernel learning model for cancer subtype discovery and essential gene identification. *Frontiers in genetics*, 12:647141, 2021. doi: 10.3389/fgene.2021.647141.

[6] Herrmann, Benjamin, Baddoo, Peter J, Semaan, Richard, Brunton, Steven L, and McKeon, Beverley J. Data-driven resolvent analysis. *Journal of Fluid Mechanics*, 918:A10, 2021. doi:10.1017/jfm.2021.337.

[7] Schmid, Peter J. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010. doi:10.1017/S0022112010001217.

[8] Schönemann, Peter H. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966. doi:10.1007/BF02289451.

[9] Speicher, Nora K and Pfeifer, Nico. Integrating different data types by regularized unsupervised multiple kernel learning with application to cancer subtype discovery. *Bioinformatics*, 31(12):i268–i275, 2015. doi:10.1093/bioinformatics/btv244.

[10] Tu, Jonathan H. *Dynamic mode decomposition: Theory and applications*. PhD thesis, Princeton University, 2013.

[11] Viguerie, Alex, Grave, Malú, Barros, Gabriel F, Lorenzo, Guillermo, Reali, Alessandro, and Coutinho, Alvaro LGA. Data-driven simulation of fisher–kolmogorov tumor growth models using dynamic mode decomposition. *Journal of Biomechanical Engineering*, 144(12):121001, 2022. doi: 10.1115/1.4054925.

Comparison of Exact Dynamic Mode Decomposition and Linear and Nonlinear Disambiguation Optimization for Data-Driven Modeling of Tumor Growth

Sakine Esmaili[(1)] [15] and Najme Mansoujian[(1)]

[(1)] Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, P.O. Box 14115-134, Tehran, Iran

Communicated by: . . . . . .

Received: . . . Accepted: . . .

Abstract: In this paper, the performances of two data-driven methods, Exact Dynamic Mode Decomposition ($exDMD$) and Linear and Nonlinear Disambiguation Optimization ($LANDO$), are compared using data generated from a dynamical system modeling tumor growth. The tumor growth dataset consists of 33 features. Each feature represents the scaled density of proliferating cells at different spatial locations within the tumor, while the samples represent the temporal evolution of cell density at different time instances. The $exDMD$ method adopts a linear approach and performs effectively in the analysis and modeling of systems with approximately linear behavior; however, it may encounter limitations when dealing with nonlinear dynamics. In contrast, $LANDO$ employs kernel-based techniques and sparse dictionaries, enabling the effective identification and separation of the linear and nonlinear components of a dynamical system. In this study, after presenting the conceptual and mathematical foundations of both methods, their performance is evaluated by implementing them on data generated from a mathematical model of tumor growth in the MATLAB environment. Furthermore, different levels of noise are added to the data, and the resulting performance is compared. The comparative results demonstrate that $LANDO$ achieves higher accuracy in modeling and representing the nonlinear behavior of the system than $exDMD$. Moreover, even the predictions obtained using the linear operator identified by $LANDO$ are more accurate than those produced by $exDMD$. These findings may contribute to improved analysis and prediction of the dynamics of diseases such as cancer.

Keywords: Exact Dynamic Mode Decomposition ($exDMD$), Singular Value Decomposition, Kernel, Linear and Nonlinear Disambiguation Optimization ($LANDO$).

©2024 Shahid Chamran University of Ahvaz, Ahvaz, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-NonComertial 4.0 International (CC BY-NC 4.0 license) (http://creativecommons.org/licenses/by-nc/4.0/).

[15]Corresponding author
E-mail addresses: (S.Esmaili) Sakine.esmaili@modares.ac.ir