مجله کنترل
ISSN (print) 2008-8345
ISSN (online) 2538-3752
مقاله پژوهشی
جلد۱۶، شماره ۱، بهار ۱۴۰۱، صفحه ۶۲-۴۹

توسعه معیار جامع سنجش کیفیت مدل‌های حاصل از شناسایی سیستم‌های هیبرید غیرخطی

احمد مداری[۱]، حمیدرضا مومنی[۲]

[۱] دانشجوی دکتری مهندسی برق-کنترل ، گروه کنترل، دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران، ایران
ahmad.madary@modares.ac.ir
[۲] پروفسور، گروه کنترل، دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران، ایران
momeni_h@modares.ac.ir

دریافت: ۱۴۰۰/۰۱/۲۹ ویرایش: ۱۴۰۰/۰۳/۱۲ پذیرش: ۱۴۰۰/۰۴/۰۱

چکیده: در این مقاله، یک معیار جامع سنجش کیفیت به منظور بررسی عملکرد مدل‌های شناسایی شده از سیستم‌های هیبرید غیرخطی با روش‌های مبتنی بر رگرسیون بردار پشتیبان، توسعه داده شده است. معیار سنجش کیفیت پیشنهادی در بردارنده تمامی فاکتورهایی است که کیفیت مدل شناسایی شده را تحت تاثیر قرار می‌دهد. این فاکتورها عبارتند از: خطای شناسایی، کیفیت سیگنال سوئیچ و پیچیدگی مدل. با استفاده از معیار کیفیت توسعه داده شده، می‌توان پاسخ‌های حاصل از شناسایی سیستم هیبرید را مقایسه کرده و بهترین مدل بدست آمده که دارای پیچیدگی معقول بوده، خطای شناسایی مناسب داشته و کیفیت سیگنال سوئیچ آن مطلوب است را انتخاب نمود. این معیار کیفیت با لحاظ کردن اصل تیغ اوکام، از انتخاب مدل‌های بسیار پیچیده جلوگیری می‌کند. همچنین امکان مقایسه تاثیر توابع کرنل متفاوت بر مدل شناسایی شده را با در نظر گرفتن فاکتورهای ذکر شده، فراهم می‌کند.

کلمات کلیدی: شناسایی سیستم، سیستم هیبرید غیرخطی، کیفیت شناسایی، اصل اوکام.

Development of A Comprehensive Quality Measure for Identification of Nonlinear Hybrid Systems

Ahmad Madary, Hamidreza Momeni

Abstract: In this study, a comprehensive quality measure criterion is developed to evaluate the performance of the identified models for nonlinear hybrid systems using support vector regression-based techniques. The proposed quality measure criterion includes all the factors that affect the quality of the identified models, namely identification error, quality of the switching signal, and model complexity. Using the proposed criterion, the resulting models of hybrid systems identification can be efficiently compared and the best model with acceptable complexity, tolerable identification error, and desirable switching signal quality will be selected. This quality measure criterion prevents selecting the complex models relying on the Occam's Razor theorem. Besides, it provides the possibility of comparing the effects of different kernel functions on the identified models considering the aforementioned factors.

Keywords: System identification, Nonlinear hybrid system, Identification quality, Occam's Razor.

۱- مقدمه

در دهه‌های اخیر، استفاده از مدل‌های هیبرید برای بیان سیستم‌ها رونق فراوانی یافته است و از این رو، مقوله شناسایی سیستم‌های هیبرید اهمیت ویژه‌ای پیدا کرده است. تاکنون روش‌های متعددی برای شناسایی سیستم‌های هیبرید ارائه شده که بیشتر آنها در زمینه سیستم‌های هیبرید خطی بوده است. این روش‌ها را می‌توان به چندین دسته اصلی شامل روش‌های مبتنی بر خوشه بندی[1] [۱،۲]، روش های بیزین [۳،۴]، روش‌های برنامه‌ریزی ترکیبی صحیح[2] [۵،۶]، روش خطای کراندار [۷،۸]، روش جبری [۹،۱۰]، و روش های مبتنی بر رگرسیون بردار پشتیبان [۱۱] طبقه بندی کرد. علاوه بر طبقه‌بندی ذکر شده، روش‌های دیگری مانند بهینه‌سازی مجموع نُرم[3] [۱۲] و روش‌های کرنل با استفاده از الگوریتم ترکیبی پایدار spline[4] [۱۳] نیز برای شناسایی سیستم های هیبرید خطی توسعه داده شده اند. به منظور مشاهده جزئیات بیشتر در خصوص روش‌های بیان شده در زمینه شناسایی سیستم‌های هیبرید خطی میتوان به [۱۴،۱۵،۱۶] مراجعه کرد.

در مقایسه با سیستم های هیبرید خطی، تحقیقات به مراتب کمتری در زمینه شناسایی سیستم های هیبرید غیرخطی انجام شده است. در مرجع [۱۷]، روش رگرسوین بردار پشتیبان به منظور شناسایی سیستم‌های هیبرید غیرخطی مورد استفاده قرار گرفته است. محققان در [۱۸] تلاش کرده اند که روش [۱۷] را برای شناسایی سیستم‌های هیبرید غیرخطی با تعداد زیاد نقاط اطلاعات مورد استفاده قرار دهند. آنان برای این کار از روش های کاهش ابعاد ماتریس کرنل مانند انتخاب بردار ویژگی[5] و تحلیل اجزاء اصلی ماتریس کرنل[6] به عنوان یک پیش پردازش استفاده کرده تا به این صورت تعداد نقاط اطلاعات را کاهش داده و یک زیرمجموعه از اطلاعات اصلی را تشکیل دهند. در مرحله بعد و با استفاده از روش [۱۷]، یک سیستم هیبرید غیرخطی برای این زیرمجموعه از اطلاعات شناسایی شده است. سپس با استفاده از این سیستم شناسایی شده، هر یک از نقاط اطلاعات مجموعه اصلی به یک زیر سیستم تخصیص داده شده و در نهایت هر یک از زیر سیستم ها توسط یک تخمین زننده متداول غیرخطی شناسایی شده است. این روش برای یک سیستم هیبرید غیرخطی با دو زیرسیستم (یک زیرسیستم خطی و یک زیرسیستم غیرخطی) و با استفاده یک تابع کرنل خطی و یک تابع کرنل گاوسی مورد استفاده قرار گرفته است. هرچند، از آنجایی که در مدل‌های تکه ای[7]، هر یک از زیرسیستم ها تنها در یک ناحیه خاص از فضای ورودی فعال است، روش ارائه شده ممکن است منجر به تولید پاسخ های زیربهینه شده و بردارهای پشتیبانی خارج از ناحیه فعال زیرسیستم را به آن تخصیص دهد. لازم به ذکر است که نحوه به دست آوردن نمایش تنک سیستم در این روش باید مورد بررسی قرار بگیرد [۱۹]. همچنین، عملکرد این روش در مواجهه با سیستم هایی که تمامی زیرسیستم های آن غیرخطی بوده و یا توابع کرنل استفاده شده شباهت زیادی با هم داشته باشند، باید مورد مطالعه قرار گیرد.

در مرجع [۲۰] یک روش بهینه‌سازی تنک بر پایه نرم $\ell_0 - \ell_1$ برای شناسایی سیستم هیبرید غیرخطی ارائه شده است. این شیوه در [۲۱] با روش‌های رگرسیون بردار پشتیبان توابع کرنل تلفیق شده و برای سیستم‌های هیبرید غیرخطی پیشنهاد شده است. مرجع [۲۲] یک روش تصادفی برای شناسایی سیستم‌های هیبرید غیرخطی ارائه داده است. در این روش، سیستم غیرخطی به صورت ترکیبی خطی از توابعی غیرخطی، به صورت خاص بسط توابع چندجمله ای با حداکثر درجه معلوم، در نظر گرفته شده است. در این شیوه از دو تابع توزیع احتمال به منظور تخصیص هر نقطه به یک زیرسیستم و نیز انتخاب ساختار مدل برای مدل‌های محلی استفاده می‌کند. یکی از فرضیات استفاده شده در این روش این است که سیگنال سوئیچ (یا تخمینی اولیه از آن) از پیش مشخص است. به این ترتیب، مجموعه نقاط اطلاعات به چندین زیرمجموعه متناظر با مودهای مختلف NARX[8] تقسیم بندی شده و پارامترهای هر یک از مودها تخمین زده می‌شود. در حالتی که هیچ اطلاعاتی از سیگنال سوئیچ در دسترس نباشد، زمان سوئیچ بر روی تمام بازه اطلاعات تغییر داده شده و تمامی ترکیب های ممکن از مودها و زمان‌های سوئیچ با استفاده از روش مونت کارلو ایجاد می‌شود. این روش پیشنهادی دارای چندین پارامتر بسیار مهم است که توسط کاربر انتخاب شده و عملکرد روش و همگرایی آن را تحت تاثیر قرار می‌دهند. در [۲۳]، یک روش شبیه سازی تصادفی زنجیره مارکوف–مونت کارلو برای شناسایی یک سیستم سوئیچ کننده شامل یک زیرسیستم خطی و یک زیرسیستم غیرخطی پیشنهاد شده است. محققان در [۲۴] از روش بیشینه کردن امید ریاضی (EM[9]) برای شناسایی یک دسته خاص از سیستم‌های هیبرید غیرخطی تحت عنوان سیستم سوئیچ کننده مارکوف اتورگرسیو دارای بخش خارجی (ARX[10]) یا سوئیچ شونده غیرخطی اتورگرسیو دارای بخش خارجی (SMNARX[11]) استفاده شده است. در این روش، هر دینامیک پیوسته توسط یک مونومیال با حداکثر درجه ثابت و مشخص نشان داده می‌شود. تخمین سیستم کلی شامل تخمین پارامترهای مدل‌های SMNARX، مقادیر حالت‌های پنهان و مقادیر مرتبط با تغییر حالت است که با روش بیشینه سازی امید ریاضی انجام می‌شود. در مرجع [۲۵]، مسئله شناسایی سیستم‌های هیبرید غیرخطی با استفاده از بسط توابع غیرخطی توسط توابع چندجمله‌ای بیان شده است. استفاده از بسط توابع چندجمله ای در دو تحقیق اخیر منجر به نفرین ابعادی[12] می‌شود. به

[1] Clustering techniques
[2] Mixed integer programming techniques
[3] Sum of norm optimization
[4] Hybrid stable spline algorithm
[5] Feature vector selection
[6] Principal component analysis
[7] Piece-wise models
[8] Nonlinear autoregressive exogenous
[9] Expectation maximization
[10] Autoregressive exogenous
[11] Switched mode nonlinear autoregressive exogenous
[12] Curse of dimensionality

منظور رهایی از این مشکل، یک روش دومرحله‌ای بازگشتی طراحی شده است. در مرحله اول، مدل‌های NARX با استفاده از یک مجموعه زمان سوئیچ تخمین زده می‌شود؛ در حالی که مرحله دوم سعی می‌کند که زمان سوئیچ را از طریق روش‌های تصادفی اصلاح کند. در مرجع [۲۶]، از روش بیشینه‌سازی امید ریاضی برای شناسایی سیستم سوئیچ کننده غیرخطی به فرم مدل‌های Hammerstein استفاده شده است. روش پیشنهادی دارای دو مرحله است: انتخاب مدل و شناسایی پارامترهای مدل. انتخاب مدل توسط روش EM انجام شده در حالی که پارامترهای مدل با استفاده از روش حداقل مربعات وزن دار شناسایی می‌شود.

۵-۱ نوآوری

در حالی که روش‌های مبتنی بر بسط مدل‌های چندجمله ای [۲۲،۲۵] دارای ویژگی‌های جذابی بوده و می‌توانند بازه وسیعی از مدل‌های غیرخطی را نمایندگی کنند، کماکان از نظر تعمیم‌پذیری توانایی کمتری نسبت به روش‌های مبتنی بر کرنل دارند. ایراد روش تصادفی ارائه شده در [۲۲] این است که پارامترهای کنترل کننده بهینگی و همگرایی روش توسط طراح انتخاب می‌شود و در صورت انتخاب نادرست عملکرد روش به شدت کاهش پیدا می‌کند. همچنین در این روش نیاز به دانستن سیگنال سوئیچ یا تخمینی از آن است که در صورتی که این مقدار در دسترس نباشد، پیچیدگی روش به صورت قابل ملاحظه ای افزایش می‌یابد. در نقطه مقابل، روش‌های مبتنی بر رگرسیون بردار پشتیبان دارای ویژگی تعمیم‌پذیری بسیار زیادی بوده و می‌توانند به همه توابع غیرخطی اعمال شوند. هرچند مسئله شناسایی سیستم‌های هیبرید غیرخطی با این روش منجر به یک مسئله بهینه‌سازی غیرمحدب می‌شود [۱۱]. عدم محدب بودن مسئله منجر به این می‌شود که مسئله دارای تعداد زیادی پاسخ نزدیک به بهینه باشد [۱۷]. همچنین با هر بار تکرار حل مسئله، پاسخ به دست آمده متفاوت از تکرار قبل می‌شود. به این ترتیب انتخاب بهترین پاسخ از میان پاسخ‌های گوناگون امری بسیار مهم است. برای این منظور می‌بایست تمامی فاکتورهای مهم و تاثیرگذار در شناسایی سیستم‌های هیبرید غیرخطی در نظر گرفته شود. این فاکتورها عبارتند از: خطای شناسایی، درصد تخصیص صحیح مود (شناسایی سیگنال سوئیچ) و پیچیدگی مدل. در روش‌های کنونی مبتنی بر رگرسیون بردار پشتیبان، تنها دو فاکتور خطای شناسایی و درصد تخصیص صحیح مود به صورت جداگانه مورد استفاده قرار می‌گیرند و معمولاً مدلی که دارای کمترین خطای برازش باشد به عنوان مدل نهایی انتخاب می‌شود. این امر می‌تواند منجر به انتخاب مدلی با پیچیدگی بیش از حد شود چراکه مدل‌های پیچیده‌تر دارای خطای برازش کمتری هستند. علاوه بر این، نوع تابع کرنل انتخاب شده به میزان زیادی در نتیجه شناسایی موثر است. با توجه به مطالب گفته شده، به منظور انتخاب بهترین مدل حاصل از شناسایی، باید تمامی فاکتورهای ذکر شده در قالب یک معیار جامع سنجش کیفیت مدل شناسایی شده لحاظ گردد تا به این وسیله بتوان مقایسه جامعی میان پاسخ‌های مختلف حاصل از شناسایی و نیز تاثیر عواملی مانند نوع تابع کرنل و مقادیر مختلف انتخاب شده برای آن، انجام داد.

در این تحقیق، یک معیار جامع سنجش کیفیت برای بررسی و مقایسه عملکرد مدل‌های حاصل از شناسایی سیستم‌های هیبرید غیرخطی با استفاده از روش‌های مبتنی بر رگرسیون بردار پشتیبان ارائه شده است. معیار ارائه شده با در نظر گرفتن تمامی فاکتورهای مهم در مسئله شناسایی، امکان مقایسه مدل‌ها و پاسخ‌های مختلف حاصل از مسئله شناسایی را فراهم می‌کند. به علاوه، این معیار جامع با لحاظ کردن اصل تیغ اوکام[1]، می‌تواند بهترین مدلی را که دارای پیچیدگی معقول بوده و دقت شناسایی مطلوبی از هر دو جنبه خطای شناسایی و دقت سیگنال سوئیچ دارد را انتخاب نماید. همچنین با استفاده از این معیار، می‌توان اثرات توابع هسته مختلف را بر مدل شناسایی شده بررسی کرده و بهترین تابع هسته به منظور شناسایی سیستم هیبرید غیرخطی را انتخاب کرد.

۲- بیان مسئله

سیستم هیبرید به فرم سیستم‌های سوئیچ شونده غیرخطی اتورگرسیو دارای بخش خارجی را می‌توان به صورت زیر نمایش داد:

$$y_i = f_{\lambda_i}(\mathbf{X}_i) + e_i \qquad (۱)$$

که در آن $\mathbf{X}_i = [y_{i-1} \cdots y_{i-n_a} \cdots u_{i-1-n_k} \cdots u_{i-n_b-n_k}]$ حالت پیوسته متشکل از n_a و n_b نمونه از ورودی u و خروجی y دارای لنگی[2]، n_k نشان دهنده تعداد نمونه‌های تاخیر یافته و e_i نویز اندازه گیری است. متغیر $\lambda_i \in \{1. \ldots . n\}$ نشان دهنده مود گسسته بوده که مشخص می‌کند کدام یک از n زیرسیستم f_j در هر لحظه فعال است. چنانچه توابع f_j غیرخطی باشند، سیستم حاصل یک سیستم سوئیچ شونده غیرخطی ARX (SNARX) خواهد بود. در روش رگرسیون بردار پشتیبان، هر یک از مدل‌های غیرخطی به‌صورت بسطی از توابع هسته به فرم زیر در نظر گرفته می‌شوند [۲۳]:

$$f_j(\mathbf{X}; \boldsymbol{\alpha}_j, b_j) = \sum_{i=1}^{N} \alpha_{ij} k_j(\mathbf{X}_i, \mathbf{X}) + b_j, \qquad (۲)$$

که در این رابطه وزن‌های $\boldsymbol{\alpha_j} = [\alpha_{1j} \cdots \alpha_{Nj}]^T$ و عبارت بایاس b_j پارامترهای $\mathbf{j}$–امین زیرمدل بوده و $k_j(.)$ تابع کرنل است. مساله شناسایی سیستم‌های هیبرید غیرخطی، مساله تعیین پارامترهای زیرمدل‌های غیرخطی $\{f_j\}_{j=1}^{n}$ (وزن‌های $\boldsymbol{\alpha}_j$ و عبارت بایاس b_j) و سیگنال سوئیچ $\lambda_i \epsilon \{1, \cdots, n\}$، با توجه به داده‌های آموزش $\mathcal{D}$ می‌باشد. این پارامترها از طریق حل مسئله ای به فرم زیر بدست می‌آیند [۱۹]:

[1] Occam's razor

[2] Lagged input and output

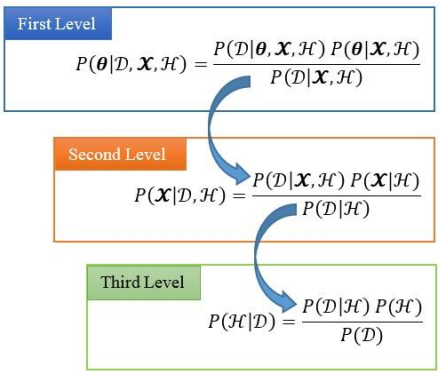
$$\underset{\{\alpha_j\},\{b_j\}}{minimize}\ \frac{1}{n}\sum_{j=1}^{n} R(\boldsymbol{\alpha}_j) + \frac{C}{N}\sum_{i=1}^{N}\left(\min_{j=1,\dots n}\ell\left(y_i-\sum_{k=1}^{N}\alpha_{kj}\,k_j(\boldsymbol{x}_i,\boldsymbol{x}_k)+b_j\right)\right) \quad (۳)$$

در این رابطه، پارامتر C وظیفه تنظیم پیچیدگی مدل و خطای برازش اطلاعات را بر عهده دارد. همان‌طور که پیشتر بیان شد، یک معیار جامع سنجش کیفیت مدل شناسایی شده برای سیستم‌های هیبرید غیرخطی می‌بایست دربردارنده فاکتورهای *خطای شناسایی*، *دقت تخمین سیگنال سوئیچ*، *پیچیدگی مدل*، و *تاثیر تابع کرنل انتخاب شده* باشد.

دقت برازش اطلاعات یا میزان خطای شناسایی توسط پارامترهای مدل یعنی (وزن‌های $\boldsymbol{\alpha}_j$ و عبارت بایاس b_j) تعیین می‌شود. همچنین در سیستم‌های هیبرید، دقت تخمین سیگنال سوئیچ را می‌توان به صورت میزان نقاط اطلاعاتی که به صورت صحیح به هر زیر سیستم تخصیص داده شده، بیان کرد. به علاوه، میزان پیچیدگی مدل توسط پارامتر C کنترل شده و اثر تابع کرنل انتخاب شده در ماتریس کرنل مشاهده می‌شود. با توجه به این مطالب، معیار جامع سنجش کیفیت مدل شناسایی شده برای سیستم‌های هیبرید غیرخطی می‌بایست در بردارنده تمامی موارد ذکر شده باشد.

به منظور به دست آوردن چنین معیاری، ابتدا مسئله شناسایی سیستم‌های هیبرید غیرخطی در چارچوب بیزین بیان می‌گردد و معیار جامع سنجش کیفیت مدل شناسایی برای این مسئله بدست می‌آید. در انتها ارتباط میان معیار بدست آمده و مدل رگرسیون بردار پشتیبان (۳) توضیح داده می‌شود.

مراحل دستیابی به معیار جامع سنجش کیفیت در شکل زیر نشان داده شده است:



شکل ۱: مراحل بدست آوردن معیار جامع سنجش کیفیت

در سطح اول، بردار پارامترهای مدل $\boldsymbol{\theta}=[\boldsymbol{\alpha},\boldsymbol{b}]^T$ محاسبه می‌شوند که در آن $\boldsymbol{\alpha}=[\boldsymbol{\alpha}_1\ \cdots\ \boldsymbol{\alpha}_n]^T$ بردار وزن‌های مدل بوده و $\boldsymbol{b}=[b_1\ \cdots\ b_n]^T$ بردار مربوط به عبارت‌های بایاس می‌باشد (n تعداد زیرسیستم‌ها را نشان می‌دهد)؛ اثرات بردار آبَرپارامترهای مدل $\boldsymbol{\mathcal{X}}=[\boldsymbol{\mu}\ \beta]$ در سطح دوم لحاظ می‌شود. این بردار دربردارنده متغیرهای واریانس برای توزیع احتمال پیشین وزن‌ها و تخمین واریانس نویز است؛ در نهایت در سطح سوم، خانواده توابع کرنل $\mathcal{H}=\{\mathcal{H}_j|j=1,\cdots,n\}$ لحاظ می‌شوند. خانواده توابع کرنل، خانواده مدل‌هایی با ساختار متفاوت و یا مقادیر متفاوت برای پارامترهاست. به عنوان نمونه عرض تابع کرنل گاوسی و یا درجه تابع کرنل چندجمله‌ای. همانطور که در این شکل دیده می‌شود، گواهی[1] هر سطح، شباهت[2] سطح بعدی می‌باشد. شاهد احتمال در روش بیزین از طریق انتگرال‌گیری از حاصل ضرب توزیع قبلی در شباهت بر روی تمامی پارامترها بدست می‌آید [۲۷]. به این ترتیب شاهد سطح اول دربردارنده خطای برازش است، چراکه انتگرال‌گیری بر روی پارامترهای مدل انجام شده است. به همین ترتیب، شاهد سطح دوم در بردارنده اثر پیچیدگی مدل بوده و از آنجایی که در انتگرال مربوطه شاهد سطح اول (شباهت سطح دوم) هم وجود دارد، اثرات خطای مدل نیز در شاهد سطح دوم لحاظ شده است.

به این ترتیب در خروجی سطح سوم، اثرات تابع کرنل نیز افزوده شده و معیار جامع سنجش کیفیت مدل شناسایی شده به صورت شاهد سطح سوم بدست می‌آید.

در ابتدا و در سطح اول، بردار پارامترهای مدل $\boldsymbol{\theta}=[\boldsymbol{\alpha},\boldsymbol{b}]^T$ که شامل وزن‌ها و عبارات بایاس برای هر زیرسیستم می‌باشند، از طریق حداکثر نمودن احتمال پسین آن‌ها محاسبه می‌شوند. احتمال پسین مشروط پارامترهای مدل با در اختیار داشتن داده‌های آموزش شامل N نقطه اطلاعات $\mathcal{D}=\{(x_i,y_i)\}_{i=1}^{N}$ بردار آبَرپارامترها $\boldsymbol{\mathcal{X}}$ و خانواده توابع کرنل $\boldsymbol{\mathcal{H}}$ بر اساس قانون بیز به صورت زیر محاسبه می‌شوند.

$$P(\boldsymbol{\alpha},\mathbf{b}|\mathcal{D},\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}}) = \frac{P(\mathcal{D}|\boldsymbol{\alpha},\boldsymbol{b},\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}})P(\boldsymbol{\alpha},\boldsymbol{b}|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}})}{P(\mathcal{D}|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}})}, \quad (۴)$$

که در آن $P(\boldsymbol{\alpha},\boldsymbol{b}|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}})$ توزیع احتمال پیشین پارامترهای مدل است و عبارت $P(\mathcal{D}|\boldsymbol{\alpha},\boldsymbol{b},\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}})$ "*مشابهت*" نقاط اطلاعات را نشان می‌دهد. مخرج کسر معادله (۴)، گواهی آبَرپارامترها نامیده شده و به علت اینکه تابعی از پارامترهای مدل نیست، [۲۸] معمولاً در محاسبه پارامترهای مدل نادیده گرفته می‌شود.

به منظور محاسبه توزیع پیشین پارامترهای مدل، فرض می‌گردد که پارامترهای هر زیرسیستم مستقل از پارامترهای سایر زیرسیستم‌هاست. همچنین، وزن‌های $\boldsymbol{\alpha}_j$ و عبارات بایاس b_j برای هر زیرسیستم مستقل می‌باشند. از این رو، احتمال شرطی توزیع پیشین روی پارامترهای مدل را می‌توان به صورت زیر نوشت:

$$P(\boldsymbol{\alpha},\mathbf{b}|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}}) = \prod_{j=1}^{n} P(\boldsymbol{\alpha}_j|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}})P(b_j|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}}), \quad (۵)$$

در گام بعد، فرض می‌گردد که توزیع پیشین وزن‌های $\boldsymbol{\alpha}_j$ مربوط به jامین زیرسیستم دارای توزیع نرمال با میانگین صفر و ماتریس کواریانس $\mu_j^{-1}I_N$ می‌باشد.

$$P(\boldsymbol{\alpha}_j|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{H}}) = \frac{1}{Z_{\alpha_j}}e^{-\frac{\mu_j}{2}\boldsymbol{\alpha}_j^T\boldsymbol{\alpha}_j} \quad (۶)$$

[1] Evidence
[2] Likelihood

$$Z_{\alpha_j}=\left(\frac{2\pi}{\mu_j}\right)^{\frac{N}{2}}.$$

در این مرحله می‌توان از انواع دیگر توابع پیشین مانند توزیع لاپلاس نیز استفاده نمود اما همانطور که در ادامه نشان داده خواهد شد در آن صورت نمی‌توان فرم قطعی برای گواهی به دست آورد و گواهی به دست آمده تقریبی خواهد بود [۲۸]. در رابطه (۶)، μ_j نشان دهنده میزان اطمینان پیشین ما از وزن‌هاست که در ادامه بیشتر مورد بحث و بررسی قرار خواهد گرفت.

عبارت دوم در رابطه (۵) توزیع احتمال پیشین روی عبارات بایاس است که معمولا به دلیل کمبود اطلاعات به صورت توابع پیشین ناآگاهنده[1] در نظر گرفته می‌شوند [۲۸]. با فرض توزیع پیشین ناآگاهنده برای عبارات بایاس b_j و توزیع نرمال برای وزن‌ها $\boldsymbol{\alpha}_j$، توزیع پیشین پارامترهای مدل را می‌توان به صورت زیر نوشت:

$$P(\boldsymbol{\alpha},\boldsymbol{b}|\boldsymbol{\chi},\boldsymbol{\mathcal{H}})=\frac{1}{\prod_{j=1}^{n}Z_{\alpha_j}}e^{\left(\sum_{j=1}^{n}-\frac{\mu_j}{2}\boldsymbol{\alpha}_j^T\boldsymbol{\alpha}_j\right)}. \quad (۷)$$

توزیع شرطی $P(\mathcal{D}|\boldsymbol{\alpha},\boldsymbol{b},\boldsymbol{\chi},\boldsymbol{\mathcal{H}})$ عبارت شباهت است که می‌تواند به صورت مدلی از نویز سیستم که باعث تخریب داده‌های آموزش اندازه‌گیری می‌شود، در نظر گرفته شود. به منظور نوشتن شباهت کامل داده‌ها، لازم است که در ابتدا نقاط اطلاعات به زیرسیستم مربوط به خود اختصاص داده شوند. به این منظور، اصل حداکثر شباهت[2] مورد استفاده قرار می‌گیرد [۱۱]. تخمین مود مبتنی بر اصل حداکثر شباهت برای سیستم‌های هیبرید تلاش می‌کند هر یک از نقاط اطلاعات $(\boldsymbol{x}_i,y_i)$ را به یک زیرسیستم که بیشترین احتمال را برای تولید این اطلاعات دارد، اختصاص دهد. به بیان دیگر، زیرسیستمی که شباهت داده را با توجه به مدل تخمین زده شده $\hat{f}_j$ حداکثر می‌کند، انتخاب می‌گردد. تخمین مود با استفاده از حداکثر شباهت را می‌توان به صورت زیر نمایش داد:

$$\hat{\lambda}_i=\arg\max_{j=1,\cdots,n}P\left(y_i|\boldsymbol{x}_i,\hat{f}_j\right),$$
$$P\left(y_i|\boldsymbol{x}_i,\hat{f}_j\right)=\frac{e^{-\ell\left(y_i-\hat{f}_j(\boldsymbol{x}_i)\right)}}{Z_\delta} \quad (۸)$$

در رابطه (۸)، $\ell(.)$ یک تابع تلف مناسب و Z_δ ثابت نرمال کننده احتمال است. همچنین، $\hat{f}_j$ مدل تخمین زده شده از jامین زیرسیستم است. فرض کنیم تابع مشابهت به صورت توزیع گاوسی با واریانس $\frac{1}{\beta}$ در نظر گرفته شود. این واریانس در واقع باور اولیه ما در ارتباط با واریانس نویز سیستم می‌باشد. عبارت$y_i-\hat{f}_j(\boldsymbol{x}_i)$ در رابطه (۸) خطای پیش بینی است و $P(y_i|\boldsymbol{x}_i,\hat{f}_j(\boldsymbol{x}_i))$ تابع چگالی احتمال خطای پیش بینی است [۲۹]. یک فرض رایج این است که خطاهای پیش بینی مستقل از یکدیگر هستند [۲۹]. با استفاده از این فرض، شباهت کامل اطلاعات را می‌توان به فرم زیر نوشت:

$$P(\mathcal{D}|\boldsymbol{\alpha},\boldsymbol{b},\boldsymbol{\chi},\boldsymbol{\mathcal{H}})=\prod_{i=1}^{N}\arg\max_{j=1,\cdots,n}P\left(y_i-\hat{f}_j(\boldsymbol{x}_i)\right)= \quad (۹)$$

$$\prod_{i=1}^{N}\arg\max_{j=1,\cdots,n}\frac{e^{-\frac{\beta}{2}\left(y_i-\hat{f}_j(\boldsymbol{x}_i)\right)^2}}{Z_\delta}$$
$$=\prod_{i=1}^{N}\frac{1}{Z_\delta}e^{\arg\max_{j=1,\cdots,n}-\frac{\beta}{2}\left(y_i-\hat{f}_j(\boldsymbol{x}_i)\right)^2},$$
$$Z_\delta=\left(\frac{2\pi}{\beta}\right).$$

احتمال پسین پارامترهای مدل را می‌توان با ترکیب نمودن توزیع پیشین پارامترها (۷) و مشابهت کامل داده‌ها (۹) به صورت زیر محاسبه نمود:

$$P(\boldsymbol{\alpha},\boldsymbol{b}|\mathcal{D},\boldsymbol{\chi},\boldsymbol{\mathcal{H}})=\frac{\prod_{i=1}^{N}Z_\delta^{-1}\prod_{j=1}^{n}Z_{\alpha_j}^{-1}e^{-\mathcal{J}_1(\boldsymbol{\alpha},\boldsymbol{b})}}{P(\mathcal{D}|\boldsymbol{\chi},\boldsymbol{\mathcal{H}})},$$
$$\mathcal{J}_1(\boldsymbol{\alpha},\boldsymbol{b})=\sum_{j=1}^{n}\frac{\mu_j}{2}\boldsymbol{\alpha}_j^T\boldsymbol{\alpha}_j \quad (۱۰)$$
$$+\frac{\beta}{2}\sum_{i=1}^{N}\arg\min_{j=1,\cdots,n}\left(y_i-\hat{f}_j(\boldsymbol{x}_i)\right)^2.$$

در این رابطه، عبارت نرمال کننده $P(\mathcal{D}|\boldsymbol{\chi},\boldsymbol{\mathcal{H}})$ گواهی اَبَرپارامترهاست که به صورت شباهت در سطح بعدی استنتاج مورد استفاده قرار خواهد گرفت. به منظور به دست آوردن پارامترهای مدل، این توزیع احتمال پسین می‌بایست حداکثر گردد که در نتیجه آن تخمین پسین پارامترها که به صورت $\boldsymbol{\alpha}^{MAP}$ و $\boldsymbol{b}^{MAP}$ نمایش داده می‌شود، حداکثر خواهد شد. حداکثر نمودن این عبارت معادل با حداقل نمودن منفی لگاریتم توزیع پسین طبق رابطه زیر است:

$$\boldsymbol{\alpha}^{MAP},\boldsymbol{b}^{MAP}:\min_{\alpha,b}\sum_{j=1}^{n}\frac{\mu_j}{2}\boldsymbol{\alpha}_j^T\boldsymbol{\alpha}_j \quad (۱۱)$$
$$+\frac{\beta}{2}\sum_{i=1}^{N}\arg\min_{j=1,\cdots,n}\left(y_i-\hat{f}_j(\boldsymbol{x}_i)\right)^2$$

با مقایسه رابطه (۳) و رابطه (۱۱)، میتوان مشاهده کرد که این دو دقیقاً مشابه یکدیگر بوده و رابطه (۱۱) در واقع دوگان رابطه (۳) در فضای بیزین است و خروجی هر دو رابطه، پارامترهای مدل شامل وزن‌ها و عبارت‌های بایاس است.

اکنون و با در دست داشتن پارامترهای مدل، زیرسیستم‌ها را می‌توان با استفاده از تخمین حداکثر شباهت (۸) که در ادامه به صورت زیر بازنویسی شده است، با حل یک مساله بهینه سازی با متغیرهای پیوسته و گسسته تخمین زد.

$$\hat{\lambda}_i=\arg\min_{j=1,\cdots,n}P\left(y_i|x_i,\hat{f}_j(.;\boldsymbol{\alpha}^{MAP},\boldsymbol{b}^{MAP})\right), \quad (۱۲)$$
$$i=1,\cdots,N.$$

رابطه (۱۱) یک مساله بهینه سازی با متغیرهای پیوسته و گسسته است که حل آن نیازمند استفاده از روش‌های برنامه‌ریزی ترکیبی صحیح است. اما مشکل اساسی در استفاده از رابطه (۱۱)، گسسته بودن این رابطه به دلیل وجود عملگر min است. برای رهایی از این مشکل، تابع حداقل با یک تقریب هموار[3] به فرم تابع حداقل سازی لگاریتم مجموع نمایی[4] یا به

[1] Non-informative prior
[2] Maximum likelihood principle
[3] Smooth
[4] Min logarithm summation of exponential

اختصار MinLSE جایگزین شده است. بر اساس این تقریب تابع MinLSE به صورت زیر تعریف می‌شود.

تعریف ۱: تابع Min LSE برای یک مجموعه از اطلاعات به صورت $\{x_j\}_{j=1}^n$ به فرم زیر تعریف می‌گردد که در آن، $\kappa > 0$ یک ضریب مقیاس[1] برای ارتقای بیشتر دقت تقریب است.

$$Min\ LSE\ (x_1, \cdots, x_n) = -\kappa^{-1} \log\left(\sum_{j=1}^{n} \exp(-\kappa x_j)\right). \quad (۱۳)$$

حداکثر اختلاف MinLSE نسبت به کمینه اصلی تنها به تعداد آرگومان‌های تابع بستگی دارد. کران پایین MinLSE در تئوری که در ادامه معرفی می‌گردد نشان داده می‌شود.

لم ۱: تقریب MinLSE تابع حداقل سازی برای یک مجموعه دارای n متغیر $\{x_j\}_{j=1}^n$ دارای کران‌های بالا و پایین به فرم زیر است:

$$\min\{x_1, \cdots, x_n\} - \kappa^{-1}\log(n) \leq Min\ LSE\ (x_1, \cdots, x_n) < \min\{x_1, \cdots, x_n\}. \quad (۱۴)$$

اثبات: عبارت $\min_{j=1,\cdots,n}\{x_j\}$ را می‌توان به صورت زیر نوشت:

$$\min_{j=1.\cdots.n}\{x_j\} = -\kappa \log\left(\exp\left(\max_{j=1.\cdots.n}\{-\kappa x_j\}\right)\right)$$

عبارت لگاریتم سمت راست دارای حد بالایی به فرم زیر است:

$$\log\left(\exp\left(\max_{j=1,\cdots,n}\{-\kappa x_j\}\right)\right) \leq \log\left(\sum_{j=1}^{n}\exp(-\kappa x_j)\right) \leq \log(n \times \exp(-\kappa x^*)) = \log n - \kappa x^*. \quad (۱۵)$$

که در این رابطه، x^*، مقدار حداقل مجموعه $\{x_1.,,,.x_n\}$است.

با ضرب نمودن سمت راست آخرین عبارت در $-\kappa^{-1}$، کران بالا و پایین در رابطه (۱۴) به دست می‌آید. □

ذکر این نکته لازم است که با توجه به رابطه اخیر، با انتخاب یک مقدار مناسب برای پارامتر κ، می‌توان کران پایین تقریب را به قدر کافی کوچک نمود. با استفاده از تابع MinLSE مساله بهینه سازی (۱۱) به صورت زیر بازنویسی می‌شود.

$$\boldsymbol{\alpha}^{MAP}, \boldsymbol{b}^{MAP}: \min_{\alpha,b} \mathcal{J}_1 = \min_{\alpha,b} \sum_{j=1}^{n} \frac{\mu_j}{2}\boldsymbol{\alpha}_j^T\boldsymbol{\alpha}_j + \frac{\beta}{2}\sum_{i=1}^{N} \underset{j=1,\cdots,n}{\text{MinLSE}}\left(\left(y_i - \hat{f}_j(\boldsymbol{x}_i)\right)^2\right) \quad (۱۶)$$

توزیع پسین پارامترهای مدل را می‌توان با استفاده از مقادیر محاسبه شده برای $\boldsymbol{\alpha}^{MAP}$ و $\boldsymbol{b}^{MAP}$ و میله‌های خطا[2] بر روی این پارامترهای پسین حداکثر خلاصه کرد. میله‌های خطا از روی خمیدگی[3] توزیع پسین محاسبه می‌شوند [۲۸]. توزیع پسین می‌تواند به صورت محلی توسط یک توزیع گاوسی به صورت زیر تقریب زده شود:

$$P(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\chi}, \boldsymbol{\mathcal{H}}) \cong P(\boldsymbol{\theta}^{MAP}|\mathcal{D}, \boldsymbol{\chi}, \boldsymbol{\mathcal{H}}) \exp\left(-\frac{1}{2}\Delta\boldsymbol{\theta}^T\Sigma\Delta\boldsymbol{\theta}\right), \quad (۱۷)$$

که در آن $\boldsymbol{\theta}^{MAP} = [\boldsymbol{\alpha}^{MAP}, \boldsymbol{b}^{MAP}]^T$ و $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}^{MAP}$. در رابطه (۱۷)، Σ ماتریس هسین $\Sigma = -\nabla^2 \log P(\boldsymbol{\alpha}, \boldsymbol{b}|\mathcal{D}, \boldsymbol{\chi}, \boldsymbol{\mathcal{H}})$ است و ماتریس کواریانس $\mathcal{J}_1$ (میله‌های خطا) برابر است با Σ^{-1} . دقت این تقریب به نوع مساله مورد بررسی وابسته است. برای عبارت درجه دوم[4] استفاده شده در این مقاله، تقریب دارای دقت بالایی است [۲۸].

پس از به دست آوردن محتمل‌ترین مقادیر برای پارامترها، تخمین مود با استفاده از رابطه (۱۲) صورت می‌پذیرد و برای هر نقطه اطلاعات یک متغیر گسسته B_{ij} به صورت زیر تعریف می‌گردد که هر نقطه اطلاعات را به یک زیرسیستم ارتباط می‌دهد.

$$B_{ij} \in \{0,1\}, \quad \forall i = 1, \cdots, N \ \ j = 1, \cdots, n, \quad s.t\ B_{ij} = 1 \quad iff\ \lambda_i = j \ \ and \ \ \sum_{j=1}^{n} B_{ij} = 1 \quad (۱۸)$$

با وارد نمودن این متغیر گسسته به رابطه (۱۱)، تابع هزینه $\mathcal{J}_1$ را می‌توان به صورت زیر بازنویسی نمود:

$$\mathcal{J}_1 = \sum_{j=1}^{n} \frac{\mu_j}{2}\boldsymbol{\alpha}_j^T\boldsymbol{\alpha}_j + \frac{\beta}{2}\sum_{i=1}^{N}\sum_{j=1}^{n} B_{ij}\left(y_i - \hat{f}_j(\boldsymbol{x}_i)\right)^2. \quad (۱۹)$$

عبارت اول در این رابطه عبارت تنظیم کننده[5] نامیده می‌شود که نشان دهنده همواری مورد انتظار از مدل حاصل می‌باشد [۴]. همچنین عبارت دوم برازندگی داده‌ها است. با مقایسه رابطه اخیر با رابطه (۳)، می‌توان مشاهده کرد که ثابت C در رابطه (۳) که وظیفه تنظیم پیچیدگی مدل و خطای برازش را بر عهده دارد، معادل $\frac{\beta}{\mu_j}$ در رابطه (۱۹) است. این تناظر میان مسئله (۱۹) و شناسایی سیستم هیبرید غیرخطی با استفاده از رگرسیون بردار پشتیبان، به معیار جامع سنجش کیفیت این امکان را می‌دهد که بتوان آن را برای سنجش کیفیت شناسایی انجام شده توسط روش رگرسیون بردار پشتیبان نیز به کار برد.

علت تفکیک ثابت C به فرم $\frac{\beta}{\mu_j}$ این است که حتی در سیستم‌های غیرهیبرید، پارامترهای مدل وابستگی زیادی به مقادیر پیشین واریانس وزن‌ها و نویز دارند چرا که بسته به مقادیر پارامترهای مدل و نسبت $\frac{\beta}{\mu_j}$ می‌توانند باعث بروز مشکلات جدی بیش برازش[6] و کم برازش[7] شوند [۳۰]. در سیستم‌های هیبرید، از آنجایی که هدف تنها سازگار نمودن مدل‌ها با داده‌ها نیست و توالی سوئیچ زنی نیز می‌بایست تخمین زده شود، این مساله اهمیت بالاتری پیدا می‌کند. مقادیر نامناسب برای μ_j، β و پارامترهای

[1] Scale factor
[2] Error bars
[3] Curvature
[4] Quadratic
[5] Regularization term
[6] Over-fitting
[7] Under-fitting

مدل ممکن است به تخمین نادرست مود منجر شود (در بدترین حالت همه داده‌ها به یک مود اختصاص داده می‌شوند). ممکن است این مساله مطرح گردد که تنها نسبت $\frac{\beta}{\mu_j}$ حائز اهمیت است. این امر در صورتی صحیح است که هدف تنها به دست آوردن پارامترهای بهترین برازش[1] باشد. مزیت جدا نمودن این دو پارامتر در این است که قابلیت به کارگیری اطلاعات سایر منابع از قبیل کران مقادیر نویز، ایجاد خواهد شد.

پس از بدست آوردن مقادیر بهینه پارامترهای مدل و لحاظ کردن تاثیر آن ها در شاهد سطح اول، در مرحله بعد اثرات عدم قطعیت این پارامترها به همراه واریانس نویز به معیار سنجش کیفیت اضافه می‌شود. این موضوع معادل لحاظ کردن میزان پیچیدگی مدل است. هدف از سطح دوم استنتاج به دست آوردن مقادیر بهینه برای آَپرپارامترهای مدل شامل واریانس وزن‌ها برای هر مدل ($\frac{1}{\mu_j}$) و واریانس پیشین نویز ($\frac{1}{\beta}$) است.

در این مرحله، توزیع پسین آَپرپارامترها با در اختیار داشتن نقاط اطلاعات و مدل با استفاده از قانون بیز بیشینه می‌شود. این توزیع احتمال پسین به صورت زیر نشان داده می‌شود:

$$P(\boldsymbol{\chi}|\mathcal{D},\mathcal{H}) = \frac{P(\mathcal{D}|\boldsymbol{\chi},\mathcal{H})P(\boldsymbol{\chi}|\mathcal{H})}{P(\mathcal{D}|\mathcal{H})}, \qquad (۲۰)$$

که در آن، $P(\boldsymbol{\chi}|\boldsymbol{\mathcal{H}})$ توزیع پیشین با داشتن اطلاعات مدل $\boldsymbol{\mathcal{H}}$ می‌باشد. از آنجایی که پیش از آموزش اطلاعات کمی در ارتباط با مقادیر بهینه آَپرپارامترها در اختیار است، توزیع پیشین آن‌ها تخت[2] فرض می‌شود (تخت در مقیاس لگاریتمی چون پارامترهای مقیاس هستند). این فرض نشان می‌دهد که هیچ یک از مقادیر آَپرپارامترها نسبت به سایرین ارجحیت نداشته و همه آن‌ها به یک اندازه محتمل هستند. برای اطلاعات بیشتر در ارتباط با توزیع پیشین می‌توان به [۲۸] مراجعه نمود. همچنین، $P(\boldsymbol{\mathcal{D}}|\boldsymbol{\mathcal{H}})$ گواهی مدل است که در سطح سوم استنتاج استفاده می‌شود.

عبارت $P(\mathcal{D}|\boldsymbol{\chi},\boldsymbol{\mathcal{H}})$ شباهت داده‌های آموزش با داشتن آَپرپارامترهای مدل و خانواده مدل‌ها $\boldsymbol{\mathcal{H}}$ است که بر اساس (۴) گواهی سطح اول استنتاج است. با فرض توزیع پیشین یکنواخت برای آَپرپارامترها، حداکثر نمودن توزیع پسین معادل با حداکثر نمودن شباهت سطح دوم بوده که همان گواهی سطح اول استنتاج می‌باشد. فرض کنیم $\boldsymbol{\theta} = [\boldsymbol{\alpha}\ \boldsymbol{b}]^T$ نشان دهنده پارامترهای مدل می‌باشد. گواهی سطح اول از طریق محاسبه انتگرال زیر روی پارامترهای مدل محاسبه میگردد:

$$P(\mathcal{D}|\boldsymbol{\chi},\boldsymbol{\mathcal{H}}) = \int P(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\chi},\boldsymbol{\mathcal{H}})P(\boldsymbol{\theta}|\boldsymbol{\chi},\boldsymbol{\mathcal{H}})d\boldsymbol{\theta}, \qquad (۲۱)$$

توزیع پسین دارای پیک‌های شدیدی در مقادیری از پارامترهای مدل که احتمال وقوع بالایی دارند، می‌باشد. بنابراین انتگرال گواهی می‌تواند با توجه به پیک انتگرال و عرض آن $\Delta\theta$ تقریب زده شود [۳۱]. بهترین شباهت در فاکتور اوکام[3] که مقدار آن کوچکتر از یک است ضرب شده و مدل $\boldsymbol{\mathcal{H}}$ را با توجه به پارامتر $\boldsymbol{\theta}$ جریمه می‌کند.

$$\underbrace{P(\mathcal{D}|\boldsymbol{\chi},\boldsymbol{\mathcal{H}})}_{Evidence} \approx \underbrace{P(\mathcal{D}|\boldsymbol{\theta}^{MAP},\boldsymbol{\chi},\boldsymbol{\mathcal{H}})}_{Best\ fit\ likelihood}\underbrace{P(\boldsymbol{\theta}^{MAP}|\boldsymbol{\chi},\boldsymbol{\mathcal{H}})\Delta\boldsymbol{\theta}}_{Occam\ factor}. \qquad (۲۲)$$

اصل اوکام بیان می‌کند: "یک مدل برای برازش داده‌ها می‌بایست به اندازه کافی پیچیده باشد" و یا به بیان دیگر، یک مدل نباید بیش از حد پیچیده[4] باشد. فاکتور اوکام به مدل‌های ساده‌تر پاداش می‌دهد. یک مدل پیچیده که دارای پارامترهای زیادی است، نسبت به مدل‌های ساده‌تر بیش‌تر جریمه می‌شود. این ضریب همچنین مدل‌هایی را که برای برازش داده‌ها نیاز دارند به خوبی تنظیم شوند، بیش‌تر جریمه می‌کند [۲۸]. به بیان دیگر، مدل‌هایی تشویق می‌شوند که به دقت کمتری[5] برای پارامترهای خود نیاز دارند. با استفاده از توزیع گاوسی تقریبی برای توزیع پسین پارامترهای مدل، گواهی را می‌توان به طریق زیر محاسبه نمود:

$$P(\mathcal{D}|\boldsymbol{\chi},\boldsymbol{\mathcal{H}}) = \prod_{j=1}^{n} Z_{\alpha_j}^{-1} \prod_{i=1}^{N} Z_{\delta}^{-1} e^{-\mathcal{J}_1(\boldsymbol{\theta}^{MAP})}(2\pi)^{\frac{n(N+1)}{2}}|\Sigma|^{-\frac{1}{2}} \qquad (۲۳)$$

که در آن، $Z_{\alpha_j} = \left(\frac{2\pi}{\mu_j}\right)^{\frac{N}{2}}$ و $Z_{\delta} = \left(\frac{2\pi}{\beta}\right)^{\frac{1}{2}}$ و Σ ماتریس هسین تابع هزینه سطح اول است. آَپرپارامترهای μ_j وظیفه کنترل پیچیدگی مدل را بر عهده دارند. مدلی با مقادیر بزرگ μ_j (واریانس کوچک توزیع پیشین وزن‌ها) داده‌ها را با یک تابع هموار برازش می‌کند در حالی که مدلی با مقادیر کوچک μ_j (آزادی زیاد در گستره پیشین مقادیر ممکن $\boldsymbol{\alpha}$) داده‌ها را با تابع هموار و پیچیده برازش می‌نماید. بر اساس اصل اوکام، این پارامتر نباید خیلی بزرگ یا خیلی کوچک انتخاب شود [۳۲]. یکی از جذاب‌ترین جنبه‌های رویکرد بیزین این است که اصل اوکام می‌تواند به صورت خودکار با انتگرال‌گیری تمامی متغیرهای نامرتبط اعمال گردد. به بیان دیگر، چارچوب بیزین به صورت خودکار مدل‌های ساده را که بدون پیچیدگی اضافی و به طور کافی داده‌ها را توضیح می‌دهند ترجیح می‌دهد [۳۲]. این ویژگی حتی در صورتی که احتمال پیشین کاملا ناآگاهنده باشد، برقرار است [۲۸].

اکنون، یک تابع هزینه برای این مرحله به صورت زیر تعریف می‌کنیم که از آن به منظور بدست آوردن گواهی سطح سوم استفاده خواهیم کرد. این تابع هزینه به صورت منفی لگاریتم (۲۳) و به صورت زیر تعریف می‌گردد:

$$\mathcal{J}_2 = -\frac{N}{2}\left(\sum_{j=1}^{n}\log\mu_j + \log\beta\right) + \frac{N-n}{2}\log 2\pi \qquad (۲۴)$$

[1] Best fit parameters
[2] flat
[3] Occam's factor
[4] Overly complex
[5] Rough precision

$$+\mathcal{J}_1(\boldsymbol{\theta}^{MAP};\boldsymbol{\mu},\beta)+\frac{1}{2}\log|\Sigma|.$$

ماتریس هسین Σ می‌تواند به صورت زیر نوشته شود:

$$\Sigma=\begin{pmatrix}M_\mu+\beta H_1 & \beta H_2\\ \beta H_2^T & \beta H_3\end{pmatrix}, \qquad (۲۵)$$

به طوری که:

$$[H_{1j}]_{ts}=\left[\frac{\partial^2\mathcal{J}_1}{\partial\alpha_{tj}\partial\alpha_{sj}}\right]$$
$$=\begin{cases}if\ t=s: & \mu_z+\beta\sum_{i=1}^{N}B_{iz}k_z(x_i,x_t)^2\\ if\ t\neq s: & \beta\sum_{i=1}^{N}B_{iz}k_z(x_i,x_t)k_z(x_i,x_s)\end{cases}$$
$$[H_{2j}]_{s1}=\left[\frac{\partial^2\mathcal{J}_1}{\partial b_j\partial\alpha_{sj}}\right]=\beta\sum_{i=1}^{N}B_{iz}k_z(x_i,x_s)$$
$$[H_{3j}]=\left[\frac{\partial^2\mathcal{J}_1}{\partial b_j\partial b_j}\right]=\beta\sum_{i=1}^{N}B_{iz},$$

همچنین M_μ یک ماتریس قطری $(nN\times nN)$ است. یکی از ویژگی‌های این ماتریس هسین این است که با توجه به فورمولاسیون ارائه شده، تنک بوده و عناصر آن ماتریس‌های قطری-بلوکی[1] هستند: $M_\mu=diag(\mu_1 I_N,\cdots,\mu_N I_N)$ و $H_1=diag(H_{11},\cdots,H_{1n})\in\mathbb{R}^{nN\times nN}$، $H_2=diag(H_{21},\cdots,H_{2n})\in\mathbb{R}^{nN\times n}$ و $H_{1j}\in\mathbb{R}^{N\times N}$، $H_{2j}\in\mathbb{R}^{N\times 1}$ و $H_3=diag(H_{31},\cdots,H_{3n})\in\mathbb{R}^{n\times n}$ برای $H_{3j}\in\mathbb{R}$ و $j=1,\cdots,n$ دترمینان ماتریس هسین می‌تواند به صورت $|\Sigma|=|M_\mu+\beta H_a||\beta H_3|$ محاسبه گردد که $H_a=(H_1-H_2H_3^{-1}H_2^T)$. به دلیل ساختار قطری-بلوکی مولفه‌های Σ، ماتریس H_a نیز یک ماتریس قطری-بلوکی است که $H_a=diag(H_{a1},\cdots,H_{an})$ می‌تواند به صورت تابعی از مولفه‌های Σ به صورت $H_{aj}=\left(H_{1j}-H_{2j}H_{3j}^{-1}H_{2j}^T\right)$ نمایش داده شود. با استفاده از این نحوه نمایش، دترمینان Σ به صورت زیر خواهد بود:

$$|\Sigma|=\prod_{j=1}^{n}\left|\mu_j I_N+\beta H_{aj}\right||\beta H_3|. \qquad (۲۶)$$

لگاریتم $|\Sigma|$ را می‌توان بر اساس مقادیر ویژه غیرصفر H_{aj} به صورت زیر نوشت:

$$\log|\Sigma|=\sum_{j=1}^{n}\left(\begin{aligned}&(N-k_j)\log\mu_j\\&+\sum_{l=1}^{k_j}\log\left(\mu_j+\beta\lambda_l(H_{aj})\right)\end{aligned}\right)\qquad (۲۷)$$
$$+n\log\beta+\sum_{t=1}^{n}\log\lambda_t(H_3),$$

در این رابطه، k_j تعداد مقادیر ویژه غیرصفر H_{aj} است که تنها تابعی از کرنل و نقاط اطلاعات آموزش می‌باشد.

اکنون و تا این مرحله، اثرات پارامترهای مدل و اَبَرپارامترها در شاخص کیفیت لحاظ شده و می‌توان شاخص جامع سنجش کیفیت مدل شناسایی شده را در مرحله سوم بدست آورد. این شاخص کیفیت برابر با توزیع احتمال پسین مدل به شرط اطلاعات است. همانظور که گفته شد، پارامترهای کلیدی متعددی که بر شناسایی سیستم‌های هیبرید تاثیرگذار بوده و نقش مهمی در کیفیت مدل شناسایی شده دارند، عبارتند از: *برازندگی داده‌ها*[2]، *پیچیدگی مدل*، و *تعداد پارامترهای تخصیص داده شده به هر زیرسیستم*.

در سیستم‌های رایج غیرهیبرید، ماشین‌های بردار پشتیبان حداقل مربعات (LS-SVM[3]) [۳۳] از دو مورد اولیه برای ارزیابی پروسه شناسایی و مقایسه مدل‌های مختلف استفاده می‌کنند. اما هدف شناسایی سیستم هیبرید، تخمین پارامترهای هر زیرسیستم و سیگنال سوئیچ به طور همزمان می‌باشد. روش‌های کنونی برای شناسایی سیستم‌های هیبرید غیرخطی می‌توانند پیچیدگی مدل را از طریق ضریب مصالحه[4] در روش‌های SVR کنترل نمایند اما قادر به بکارگیری مستقیم پیچیدگی مدل با برازندگی داده‌ها برای مقایسه مدل‌های مختلف شناسایی شده یا ساختارهای مدل نیستند. همچنین میزان داده‌های اختصاص داده شده به هر زیرسیستم می‌بایست به منظور مقایسه نتایج رویه‌های شناسایی مختلف در نظر گرفته شود. طبق اطلاعات در دسترس، یک معیار یکپارچه و همه‌جانبه برای روش‌های SVR که بتواند تمامی موارد را برای سیستم‌های هیبرید پوشش دهد، در حال حاضر موجود نیست. ضرورت دیگر برای داشتن یک شاخص اندازه‌گیری یکپارچه برای مقایسه نتایج شناسایی این است که مساله شناسایی کنونی برای سیستم‌های هیبرید غیرخطی (شامل تحقیق حاضر و [۳۵])، یک مساله بهینه سازی غیرمحدب[5] است که دارای چندین پاسخ نزدیک به بهینه[6] می‌باشد. از این رو، نه تنها انتخاب ساختارهای مدل متفاوت و حتی پارامترهای مختلف مدل برای یک زیرسیستم نمونه بر کل پروسه شناسایی تاثیرگذار است، بلکه از آنجایی که پاسخ یکتا وجود ندارد، تکرارهای مختلف برای ساختارهای مدل ثابت نیز خروجی پروسه شناسایی را تغییر می‌دهند. بنابراین در اختیار داشتن یک معیار همه جانبه برای ارزیابی کیفیت پاسخ‌ها ضروری است. این شاخص جامع سنجش کیفیت اهداف زیر را برآورده می‌نماید:

- مقایسه و انتخاب پاسخ‌های مختلف مساله شناسایی؛
- مقایسه و انتخاب ساختارهای مختلف مدل؛
- مقایسه پارامترهای مدل متفاوت.

باید توجه نمود که مقایسه مدل‌های مختلف مساله دشواری است چرا که انتخاب ساده یک مدل با توجه به بهترین برازش داده‌ها بر اساس معیاری

[1] Block-diagonal matrices
[2] Data fitness
[3] Least square support vector machines
[4] Trade-off coefficient
[5] Non-convex
[6] Near-optimal

از قبیل متوسط مربعات خطا (MSE[1]) منجر به بیش برازش می‌شود زیرا مدل‌های پیچیده‌تر همواره برازش بهتری روی داده‌ها دارند. از این رو، انتخاب بهترین مدل تنها با لحاظ نمودن معیار برازش (به عنوان مثال حداکثر شباهت) به مدل‌های over-parameterized با قابلیت تعمیم‌دهی[2] ضعیف ختم خواهد شد که در اینجا اصل اوکام اهمیت پیدا می‌کند [۲۸]. سطح سوم استنتاج همچنین ابزاری برای ارزیابی تاثیر انتخاب یک مدل مشخص برای یک زیرسیستم بر کل پروسه شناسایی، ارائه می‌نماید.

توزیع بعدی مدل $\mathcal{H}_j$ به عنوان شاخص عملکردی برای آن مدل مشخص استفاده می‌شود. با فرض یک توزیع تخت پیشین برای مدل $\mathcal{H}_j$، توزیع بعدی با شباهت $P(\mathcal{D}|\mathcal{H}_j)$ متناسب خواهد بود که گواهی مدل در سطح قبل است. این توزیع بعدی دارای فرم زیر است:

$$P(\mathcal{H}_j|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{H}_j)P(\mathcal{H}_j)}{P(\mathcal{D})} \qquad (۲۸)$$

گواهی $P(\mathcal{D}|\mathcal{H}_j)$ با انتگرال‌گیری روی همه متغیرها (آبرپارامترها $\boldsymbol{\chi}_j = [\mu_j\ \beta]$) به دست خواهد آمد:

$$P(\mathcal{D}|\mathcal{H}_j) = \int P(\mathcal{D}|\boldsymbol{\chi}_j, \mathcal{H}_j)P(\boldsymbol{\chi}_j|\mathcal{H}_j)d\boldsymbol{\chi}_j \qquad (۲۹)$$

گواهی را می‌توان به طور دقیق با یک توزیع نرمال تفکیک پذیر[3] با میله‌های خطای σ_{μ_j} و σ_β تقریب زد. این میله‌های خطا با دو بار مشتق‌گیری از تابع هزینه سطح دوم (۲۴) نسبت به μ_j و β محاسبه می‌شوند:

$$\begin{aligned}\left(\sigma_{\mu_j}\right)^2 &= \left(\frac{k_j}{2\mu_j^2} - \frac{1}{2}\sum_{l=1}^{k_j}\frac{1}{\left(\mu_j + \beta\lambda_l(H_{aj})\right)^2}\right)^{-1},\\ \left(\sigma_\beta\right)^2 &= \left(\frac{N-n}{2\beta^2} - \frac{1}{2}\sum_{j=1}^{n}\sum_{l=1}^{k_j}\frac{\lambda_l(H_{aj})^2}{\left(\mu_j + \beta\lambda_l(H_{aj})\right)^2}\right)^{-1}.\end{aligned} \qquad (۳۰)$$

با استفاده از این میله‌های خطا در رابطه (۲۹) و با فرض یک توزیع پیشین تخت، گواهی به صورت زیر قابل محاسبه است:

$$P(\mathcal{D}|\mathcal{H}_j) \approx P(\mathcal{D}|\chi_j^{MAP}, \mathcal{H}_j)2\pi\sigma_\beta\sigma_{\mu_j}, \qquad (۳۱)$$

به طوری که $P(\mathcal{D}|\chi^{MAP}, \mathcal{H}_j)$ با استفاده از محتمل‌ترین مقادیر برای آبرپارامترها که در سطح قبل به دست آمده‌اند، محاسبه می‌شود. با صرف نظر از تمامی ثوابت در روابط (۲۴) و (۳۱)، توزیع پسین برای مدل $\mathcal{H}_j$ به این صورت محاسبه می‌گردد:

$$P(\mathcal{H}_j|\mathcal{D}) \propto \sqrt{\frac{\left(\mu_j^{MAP}\right)^N \sigma_{\mu_j}^2 (\beta^{MAP})^{\sum_{i=1}^N B_{ij}}\ \sigma_\beta^2 \exp\left(-\mathcal{J}_1(\theta^{MAP}, \chi^{MAP})\right)}{\left(\mu_j^{MAP}\right)^{N-k_j}\prod_{l=1}^{k_j}\left(\mu_j^{MAP} + \beta^{MAP}\lambda_l(H_{aj})\right)\beta^{MAP}H_{3j}}}. \qquad (۳۲)$$

برای راحتی بیشتر می‌توان از لگاریتم (۳۲) به عنوان سنجه‌ای از کیفیت مدل استفاده نمود. سنجه کیفیت برای مدل $\mathcal{H}_j$ به صورت زیر محاسبه می‌شود:

$$\begin{aligned}QM(\mathcal{H}_j) = \log\sigma_{\mu_j} + \log\sigma_\beta + \frac{k_j}{2}\log\mu_j^{MAP}\\ + \frac{\zeta_j - 1}{2}\log\beta^{MAP} - \frac{1}{2}\log\zeta_j\\ -\frac{1}{2}\sum_{l=1}^{k_j}\left(\mu_j^{MAP} + \beta^{MAP}\lambda_l(H_{aj})\right)\\ -\frac{\mu_j^{MAP}}{4}\parallel\alpha_j\parallel_2^2 - \frac{\beta^{MAP}}{4}\sum_{i=1}^N\beta_{ij}\left(y_i - f_j(x_i)\right)^2.\end{aligned} \qquad (۳۳)$$

در این معادله $\zeta_j = \sum_{i=1}^N B_{ij}$، تعداد نقاط اطلاعات اختصاص داده شده به مدل $\mathcal{H}_j$ بوده و k_j تعداد مقادیر ویژه غیرصفر ماتریس کرنل مربوط به $\mathcal{H}_j$ را نشان می‌دهد.

این سنجه کیفیت یک ویژگی منحصر به فرد دارد به طوری که شامل تمام مولفه‌های مرتبط تعیین کننده کیفیت شناسایی است. این مولفه‌ها عبارتند از:

- برازندگی مدل مرتبط با $\mathcal{H}_j$: $\sum_{i=1}^N B_{ij}\left(y_i - f_j(\boldsymbol{x}_i)\right)^2$ و واریانس پیشین نویز $\frac{1}{\beta^{MAP}}$؛
- پیچیدگی مدل: عبارت تنظیم کننده $\parallel\alpha_j\parallel_2^2$ و واریانس پیشین وزن‌ها μ_j^{MP}؛
- عدم قطعیت نویز و واریانس وزن‌ها: $\log\sigma_{\mu_j}$ و $\log\sigma_\beta$؛
- تعداد نقاط اطلاعات اختصاص داده شده به مدل $\mathcal{H}_j$: ζ_j؛
- مشخصه‌های ماتریس کرنل (مقادیر ویژه) مربوط به $\mathcal{H}_j$.

معیار جامع بدست آمده را می‌توان به صورت زیر جمع‌بندی نمود: این سنجه کیفیت به مدل‌های ساده با بهترین برازش اطلاعات و بیشترین نقطه اطلاعات اختصاص داده شده پاداش می‌دهد.

از آنجایی که هر تغییری در یک $\mathcal{H}_j$ نتایج شناسایی را برای سایر مدل‌ها تغییر می‌دهد، یک سنجه کیفیت کلی برای شناسایی می‌بایست تعریف گردد که تمامی تغییرات در کل خانواده مدل‌ها $\boldsymbol{\mathcal{H}}$ در نظر گرفته شود. این سنجه کیفیت به صورت مجموع QMها برای همه مدل‌ها تعریف می‌شود:

$$QM_{Overall} = \sum_{j=1}^{n} QM(\mathcal{H}_j). \qquad (۳۴)$$

ارتباط با شاخص MDL[4] و معیار آکائیک (AIC[5]):

شاخص MDL تلاش می‌کند مدلی را انتخاب نماید که بهترین فشردگی داده را دارد؛ به بیان دیگر مدلی با تعداد پارامترهای کمتر. این شاخص می‌تواند

[1] Mean square error
[2] Generalization
[3] Separable
[4] Minimum description length
[5] Akaike's criterion

به طور خام به صورت $L(\mathcal{H})+L(\mathcal{D}|\mathcal{H})$ نوشته شود که $L(\mathcal{H})$ طول توصیف کننده مدل $\boldsymbol{\mathcal{H}}$ بر حسب بیت بوده و $L(\mathcal{D}|\mathcal{H})$ طول توصیف کننده داده‌های $\mathcal{D}$ کدگذاری شده[1] توسط $\boldsymbol{\mathcal{H}}$ (می‌تواند به صورت $\log P(\mathcal{D}|\mathcal{H})$ در نظر گرفته شود) می‌باشد [۳۴]. سنجه کیفیت QM از لگاریتم رابطه (۳۱) به دست می‌آید که بسیار مشابه شاخص MDL برای سیستم‌های غیرهیبرید رایج و معیار آکائیک است که می‌تواند به صورت تقریبی از MDL در نظر گرفته شود [۲۸].

۳- مطالعه موردی

در این بخش، سیستم هیبرید غیرخطی به فرم زیر در معرض آزمایشات شناسایی متعدد قرار می‌گیرد.

$$y_i=\begin{cases}-0.4y_{i-1}^2+0.5u_{i-1}+e_i & if \quad \lambda_i=1\\ \left(0.8-0.5e^{y_{i-1}^2}\right)y_{i-1}-y_{i-1}^2+0.9u_{i-1}+e_i & if \quad \lambda_i=2\end{cases} \qquad (۳۵)$$

در ابتدا، سیستم به مدت شش بار برای کرنل‌های مشابه $\boldsymbol{\mathcal{H}}$ با پارامترهای ثابت شناسایی می‌شود که به سبب طبیعت غیرمحدب مساله منجر به مدل‌های شناسایی شده متفاوت می‌شود. سپس سنجه کیفیت QM برای مقایسه و رتبه بندی مدل‌های حاصل و انتخاب بهترین سیستم شناسایی شده به کار می‌رود. پس از آن، عرض کرنل اول تغییر داده شده و سیستم با پارامترهای کرنل متفاوت شناسایی می‌شود و از QM برای ارزیابی تاثیرات پارامترهای کرنل متفاوت و مقایسه نتایج استفاده می‌گردد.

بررسی تاثیر ثابت C: در این بخش اثرات تغییر ثابت تنظیم کننده میان پیچیدگی مدل و خطای شناسایی مورد بررسی قرار گرفته است. برای این کار سیستم (۳۵) به ازای مقادیر مختلفی از $C_j=\frac{\beta}{\mu_j}$ مورد شناسایی قرار گرفته و مدل‌های شناسایی شده توسط معیار جامع معرفی شده، مقایسه شده‌اند. برای این کار از دو مدل گاوسین ثابت $\mathcal{H}=\{\mathcal{H}_1(0.01),\mathcal{H}_2(1)\}$ استفاده شده است. نتایج به دست آمده در جدول۱ ارائه شده است. نتایج شامل عبارت تنظیم کننده (پیچیدگی مدل)، هزینه برازش، MSE، درصد تخصیص داده صحیح، تخمین واریانس وزن‌ها و نویز (به ترتیب معکوس μ و β) و گواهی مدل یا سنجه‌های کیفیت می‌باشد. در نگاه اول و تنها با لحاظ نمودن برازندگی داده‌ها (MSE) به نظر می‌رسد که مدل به دست آمده در حالت ۶، بهترین مدل است. اما QM نشان می‌دهد مدل به دست آمده در حالت ۳ بهترین مدل است علیرغم اینکه از دیدگاه MSE در رتبه سوم قرار دارد. دلیل عمده این امر پیچیدگی مدل است: تنظیم کنندگی کلی مدل که نشان دهنده پیچیدگی مدل است، در حالت ۶ بیشتر از حالت ۳ است. این بدان معناست که مدل ششم تمایل به برازش دقیق داده‌های نویزی دارد که سبب از دست رفتن قابلیت تعمیم آن می‌شود [۲۸]. همچنین مدل ۳ دارای تخصیص داده بالاتری است. از بررسی سایر حالت‌ها نیز می‌توان به جمع بندی یکسانی دست یافت.

به منظور مقایسه زیرمدل‌ها می‌توان از QM استفاده نمود. به عنوان نمونه، $\mathcal{H}_1$ در حالت ۵ ($\mathrm{QM}_5(\mathcal{H}_1)$) بالاتر از حالت ۴ ($\mathrm{QM}_4(\mathcal{H}_1)$) است چرا که دارای تعمیم‌دهی بهتری بوده و داده‌های بیشتری را به درستی اختصاص می‌دهد. عکس این مطلب را می‌توان در ارتباط با $\mathcal{H}_2$ برای این حالت‌ها بیان نمود.

تکرارهای متفاوت برای مدل‌های یکسان: همانگونه که پیش از این ذکر گردید، مساله شناسایی سیستم‌های هیبرید به سبب طبیعت غیرمحدب آن شامل چندین راهکار نزدیک به بهینه است. از این رو، دو تکرار مختلف مساله با پارامترهای یکسان، پاسخ‌های متفاوتی تولید خواهد نمود. از این رو می‌بایست پاسخ‌های مختلف را مقایسه نمود. در این بخش، عملکرد سنجه کیفیت پیشنهادی برای این شرایط ارزیابی می‌شود. برای این منظور، سیستم زیر به دفعات شش بار مورد شناسایی قرار گرفته است.

$$y_i=\begin{cases}k\sqrt{\left(\frac{y_i-1}{k}\right)^2+\frac{u_{i-1}-y_{i-1}}{s}}+e_i & y_{i-1}\le Y\\ 1.3+\sqrt{u_{i-1}}+\exp(-y_{i-1})+e_i & y_{i-1}>Y,\end{cases} \qquad (۳۶)$$

در این روند، از دو کرنل گاوسی $\mathcal{H}=\{\mathcal{H}_1(1),\mathcal{H}_2(0.5)\}$ استفاده شده و پارامترهای تنظیم کننده پیچیدگی و خطا برابر ۱۰ در نظر گرفته شده است. نتایج شناسایی در جدول۲ نشان داده شده است.

اگر از شاخص جامع سنجش کیفیت استفاده نکرده و تنها خطای شناسایی را به عنوان معیار مقایسه انتخاب کنیم، حالت دوم به عنوان بهترین مدل انتخاب می‌شود؛ در حالی که درصد صحیح تخصیص مود که نشان دهنده کیفیت سیگنال سوئیچ شناسایی شده است، تنها ۵۹ درصد است. از طرف دیگر، اگر تنها کیفیت سیگنال سوئیچ معیار مقایسه باشد، حالت سوم با ۸۱٪ به عنوان مدل برگزیده انتخاب می‌شود؛ درحالی که از نظر میزان خطای شناسایی، پس از حالت چهارم در رتبه دوم بدترین خطای شناسایی قرار می‌گیرد. از این توضیحات می‌توان دریافت که هیچ‌یک از این دو معیار به تنهایی برای انتخاب مدل برتر کافی نیست. از طرف دیگر، صرفاً با در نظر گرفتن این دو معیار، نمی‌توان بهترین مدل را انتخاب کرد. در طرف مقابل، مدل اول دارای بیشترین شاخص کیفیت است. این امر بیان کننده این موضوع است که این مدل دارای بهترین مصالحه میان پیچیدگی، خطای شناسایی و کیفیت سیگنال سوئیچ است. دلیل برتری اندک حالت اول نسبت به حالت دوم، کیفیت بهتر سیگنال سوئیچ (میزان صحیح تخصیص مود) و پیچیدگی کمتر مدل است. همچنین این حالت به دلیل دارا بودن خطای شناسایی کمتر و نیز پیچیدگی کمتر نسبت به حالت سوم، از امتیاز بالاتری برخوردار است، علی رغم اینکه حالت سوم کیفیت سیگنال سوئیچ بهتری دارد.

[1] Encode

جدول۱: نتایج شناسایی برای شش مقدار متفاوت C

پارامترها		حالت ۱	حالت ۲	حالت ۳	حالت ۴	حالت ۵	حالت ۶
تنظیم کنندگی[1]	$\mathcal{H}_1$	0.4464	0.4360	0.4346	0.6726	0.5236	0.7994
	$\mathcal{H}_2$	0.1062	0.1670	0.1095	0.1296	0.1525	0.0963
	مجموع	0.5526	0.6030	0.5441	0.8022	0.6761	0.8957
برازندگی[2]	هزینه	0.2255	0.2277	0.2274	0.2036	0.2387	0.1697
	MSE	0.0045	0.0046	0.0045	0.0041	0.0048	0.0034
تخصیص داده‌ها[3]	$\mathcal{H}_1$	68.33	90	71.66	86.66	92.5	71.66
	$\mathcal{H}_2$	80	61.66	80	67.5	51.66	70
	کلی[4]	73	73	75	79	68	71
ابرپارامترها	$\sigma_{\alpha_1}^2$	0.0134	0.0074	0.0070	0.0094	0.0078	0.0180
	$\sigma_{\alpha_2}^2$	0.0069	0.0089	0.0095	0.0074	0.0093	0.0044
	σ_e^2	0.0108	0.0069	0.0067	0.0071	0.0068	0.0092
	C_1	1.240741	1.072464	1.044776	1.323944	1.147059	1.956522
	C_2	0.638889	1.289855	1.41791	1.042254	1.367647	0.478261
کیفیت (QM)	$\mathcal{H}_1$	83.34	109.43	95.60	113.72	119.77	86.36
	$\mathcal{H}_2$	105.18	90.38	107.81	76.68	74.99	99.66
	مجموع	188.52	199.81	203.42	190.40	194.76	186.02

جدول۲: نتایج شناسایی برای شش تکرار مختلف با مدل‌های یکسان

پارامترها		حالت ۱	حالت ۲	حالت ۳	حالت ۴	حالت ۵	حالت ۶
تنظیم کنندگی	$\mathcal{H}_1$	0.3046	0.3375	0.4569	0.6671	0.2721	0.4130
	$\mathcal{H}_2$	0.9416	0.9813	1.0603	1.2355	0.9789	0.9939
	مجموع	1.2462	1.3188	1.5172	1.9026	1.2509	1.4069
برازندگی	هزینه	0.164	0.1375	0.3146	0.4060	0.1834	0.2013
	MSE	0.0033	0.0027	0.0063	0.0081	0.0037	0.004
تخصیص داده‌ها	$\mathcal{H}_1$	64.71	43.055	66.66	15.68	50.8772	47.3684
	$\mathcal{H}_2$	100	100	100	100	100	100
	کلی۵	77	59	81	59	72	70
کیفیت (QM)	$\mathcal{H}_1$	64.7183	58.1322	59.4812	15.6849	54.4683	41.3730
	$\mathcal{H}_2$	96.9853	103.2244	75.8227	103.1039	106.0259	108.5466
	مجموع	161.7036	161.3566	135.3041	118.7888	160.4942	149.9196

لازم به ذکر است تابع کرنل دوم به دلیل دارا بودن عرض کمتر نسبت به تابع کرنل اول، توانایی بیشتری در تخصیص صحیح نقاط دارد. این امر از درصد صحیح نقاط تخصیص داده شده توسط آن قابل تشخیص است. اما در عین حال، به دلیل عرض کمتر تابع کرنل، پیچیدگی آن بیشتر از کرنل اول است. به این معنی که این تابع از عمومیت کمتری برخوردار بوده و به احتمال بسیار زیاد نویز را نیز برازش می‌کند.

پارامترهای متفاوت برای مدل‌ها: در این حالت، سنجه کیفیت (۳۳) برای رتبه بندی مدل‌های مختلف به کار گرفته می‌شود. این بار سیستم (۳۵) با استفاده از چهار کرنل گاوسی با پارامترهای مختلف برای مدل $\mathcal{H}_1$ ([0.5 0.1 0.05 0.01]) شناسایی می‌شود در حالی که مدل $\mathcal{H}_2$ دارای پارامترهای ثابت برابر با ۱ می‌باشد. نتایج حاصل در *جدول۱* ارائه شده است. با توجه به جدول، تمامی مدل‌ها دارای شاخص MSE یکسانی

[1] Regularization
[2] Fitness
[3] Data assignment
[4] Overall
[5] Overall

هستند. برای حالت‌های ۲ و ۳ علیرغم دارا بودن MSE و تخصیص داده تقریبا مشابه، QM_3 از QM_2 بزرگ‌تر است که دلیل آن پیچیدگی کمتر و قابلیت تعمیم دهی بالاتر این مدل است. شناسایی در حالت ۴ تا حدی به خاطر تخصیص داده صحیح بیشتر و عمدتا به دلیل پیچیدگی کمتر مدل، از کیفیت بالاتری نسبت به حالت ۱ برخوردار است. این مطلب را می‌توان در جدول۱ مشاهده نمود.

جدول۱: نتایج شناسایی برای ۴ حالت مختلف برای $\mathcal{H}_1$

پارامترها		حالت ۱	حالت ۲	حالت ۳	حالت ۴
		$\mathcal{H}_1(0.01)$	$\mathcal{H}_1(0.05)$	$\mathcal{H}_1(0.1)$	$\mathcal{H}_1(0.5)$
تنظیم کنندگی	$\mathcal{H}_1$	0.5134	0.2231	0.1546	0.2102
	$\mathcal{H}_2$	0.2431	0.1758	0.1937	0.2237
	جمع	0.7565	0.3989	0.3483	0.4339
برازندگی	هزینه	0.2907	0.2883	0.2973	0.29
	MSE	0.0058	0.0058	0.0059	0.0058
تخصیص داده‌ها	$\mathcal{H}_1$	65	92.5	85	77.5
	$\mathcal{H}_2$	60	68.33	73.33	88.33
	کلی	62	78	78	84
ابرپارامترها	$\sigma_{\alpha_1}^2$	0.0176	0.0077	0.0056	0.0112
	$\sigma_{\alpha_2}^2$	0.0044	0.0105	0.0100	0.0071
	σ_e^2	0.0091	0.0092	0.0073	0.0088
کیفیت (QM)	$\mathcal{H}_1$	83.20	104.91	99.84	77.34
	$\mathcal{H}_2$	100.60	92.53	108.20	131.29
	جمع	188.82	197.45	208.04	208.63

حالت ۱ ($\mathcal{H}_1(0.01)$) تمایل بیشتری به برازش داده‌های نویز نسبت به حالت ۴ دارد. با این حال باید ذکر نمود برخی از داده‌های نویز ناگزیر بر مدل منطبق خواهند شد چرا که برخی از مولفه‌های نویز از داده‌های واقعی قابل تشخیص نیستند. دو حالت ۳ و ۴ کیفیت یکسانی دارند. علیرغم اینکه حالت ۴ داده صحیح بیشتری اختصاص داده است، از آنجایی که از عمومیت کمتری نسبت به حالت ۳ برخوردار است، نمی‌تواند به صورت "بسیار بهتر[1]" رتبه بندی شود. همچنین می‌توان در جدول۱ مشاهده نمود که حالت ۳ ($\mathcal{H}_1(0.1)$) عملکرد بهتری نسبت به حالت ۴ ($\mathcal{H}_1(0.5)$) دارد. این مطلب توسط *جدول۱* که در آن $QM_3(\mathcal{H}_1)$ بزرگتر از $QM_4(\mathcal{H}_1)$ است، تایید می‌شود.

۸- جمع‌بندی

در این مقاله، یک معیار جامع سنجش کیفیت مدل شناسایی شده برای سیستم‌های هیبرید غیرخطی توسعه داده شده و عملکرد این معیار در انتخاب مدلی که دارای سطح معقولی از پیچیدگی بوده و در عین حال از دقت مناسبی از نظر خطای شناسایی و کیفیت سیگنال سوئیچ برخوردار باشد، بررسی شده است. نتایج این بررسی موید این موضوع است که معیار جامع توسعه داده شده، تمامی فاکتورهای اساسی در مسئله شناسایی سیستم‌های هیبرید غیرخطی را در بر داشته و همواره به مدل‌هایی که دارای بهترین مصالحه میان پیچیدگی، خطای شناسایی و کیفیت سیگنال سوئیچ باشند، امتیاز بیشتری می‌دهد. در این شاخص سنجش کیفیت، معیار پیچیدگی مدل با لحاظ کردن اصل تیغ اوکام بررسی می‌شود. همچنین این شاخص به طراح این امکان را می‌دهد که مدل‌های بدست آمده ازکرنل‌های مختلف را به هم مقایسه کرده و از این طریق بهترین تابع کرنل و بهترین پارامترهای آن را انتخاب نماید.

تحقیقات آینده

تحقیقات آینده در این زمینه به تدوین یک چارچوب شناسایی مبتنی بر بیزین به منظور شناسایی سیستم‌های هیبرید غیرخطی اختصاص دارد تا از این طریق بتوان علاوه بر بدست آوردن تخمین‌های احتمالاتی، اصل تیغ اوکام را در مرحله شناسایی نیز وارد کرد. به این صورت، مدل‌های حاصل از شناسایی دارای بهترین مصالحه میان پیچیدگی و خطای شناسایی خواهند بود. همچنین تلاش خواهد شد که چارچوب بیزین پیشنهادی را به نحوی تغییر داد که بتوان آن را برای تعداد نقاط زیاد اطلاعات نیز به کار برد.

مراجع

[1] G. Ferrari-Trecate, M. Muselli, D. Liberati, M. Morari, "A clustering technique for the identification of piecewise affine systems," Automatica, vol. 39, no. 2, pp. 205–217, 2003.

[2] H. Nakada, K. Takaba, T. Katayama, "Identification of piecewise affine systems based on statistical clustering technique," Automatica, vol. 41,no. 5, pp. 905–913, 2005.

[3] A.L. Juloski, S. Weiland, W.P.M.H Heemels, "A Bayesian approach to identification of hybrid systems," IEEE Transactions on Automatic Control, vol. 50, no. 10, pp. 1520–1533, 2005.

[4] Y. Lu, S. Khatibisepehr, B. Huang, "A variational Bayesian approach to identification of switched ARX models," in IEEE 53rd Annual Conference on Decision and Control(CDC), 2014, pp.2542–2547.

[5] J. Roll, A. Bemporad, L. Ljung, "Identification of piecewise affine systems via mixed-integer

[1] Significantly better

[18] G. Bloch and F. Lauer, "Reduced-size kernel models for nonlinear hybrid system identification," IEEE Transactions on Neural Networks, vol. 22,no. 12, pp. 2398–2405, 2011.

[19] F. Lauer, G. Bloch, R. Vidal, "Nonlinear hybrid system identification with kernel models," in49thIEEEConferenceonDecisionandControl,CDC2010, 2010, pp.696–701.

[20] L. Bako, K. Boukharouba, S. Lecoeuche, "Anℓ0-ℓ1norm based optimization procedure for the identification of switched nonlinear systems," in 49th IEEE Conference on Decision and Control (CDC), 2010,pp.4467–4472.

[21] V. L. Le, F. Lauer, L. Bako, G. Bloch, "Learning nonlinear hybridsystems: from sparse optimization to support vector regression," in Proceedings of the 16th international conference on Hybrid systems: computation and control, 2013, pp.33–42.

[22] F. Bianchi, M. Prandini, L. Piroddi, "A randomized approach to switched nonlinear systems identification," IFAC Papers On Line, vol. 51, no. 15,pp. 281–286, 2018.

[23] A. Scampicchio, A. Giaretta, G. Pillonetto, "Nonlinear Hybrid Systems Identification using Kernel-Based Techniques," IFAC Papers On Line, vol.51, no. 15, pp. 269–274, 2018.

[24] A. Brusaferri, M. Matteucci, A. Spinelli,"Estimation of Switched Markov Polynomial NARX models, "arXivpreprintarXiv: 2009.14073,2020.

[25] F. Bianchi, M. Prandini, L. Piroddi, "A randomized two-stage iterative method for switched nonlinear systems identification," Nonlinear Analysis: Hybrid Systems, vol. 35, pp. 100818, 2020.

[26] C. Xiujun, H. Hongwei, W. Lin, X. Zhengqing, "Identification of switched nonlinear systems based on EM algorithm," In 39th Chinese Control Conference (CCC), pp.1337-1342, 2020.

[27] J. Lunze, F. Lamnabhi-Lagarrigue, Handbook of hybrid systems control: Theory, tools, applications, Cambridge University Press, 2009.

[28] David. JC. MacKay, "Bayesian interpolation," Neural computation, vol.4, no. 3, pp. 415–447, 1992.

[29] L. Ljung, System identification, in: Signal analysis and prediction, Springer, 1998, pp. 163–173.

[30] S. S. Keerthi, C.-J. Lin, Asymptotic behaviors of support vector machines with Gaussian kernel, Neural computation 15 (7) (2003) 1667–1689.

[31] D. J. MacKay, Probable networks and plausible predictions: a review of practical Bayesian methods for supervised neural networks, Network: programming," Automatica, vol. 40, no. 1,pp. 37–50, 2004.

[6] A. Bemporad, J. Roll, L. Ljung, "Identification of hybrid systems via mixed-integer programming," in Proceedings of the 40th IEEE Conference on Decision and Control, 2001, pp.786–792.

[7] A. Bemporad, A. Garulli, S. Paoletti, A. Vicino, "A bounded-error approach to piecewise affine system identification," IEEE Transactions on Automatic Control, vol. 50, no. 10, pp. 1567–1580, 2005.

[8] A. Bemporad, A. Garulli, S. Paoletti, A. Vicino, "A greedy approach to identification of piecewise affine models," in International Workshop on Hybrid Systems: Computation and Control, 2003, pp.97–112.

[9] Y. Ma, R. Vidal, "Identification of deterministic switched ARX systems via identification of algebraic varieties," in International Workshop on Hybrid Systems: Computation and Control, 2005, pp.449–465.

[10] R. Vidal, S. Soatto, Y. Ma, S. Sastry, "An algebraic geometric approach to the identification of a class of linear hybrid systems," in 42nd IEEE Conference on Decision and Control, 2003, pp.167–172.

[11] F. Lauer, "From support vector machines to hybrid system identification," Ph.D. dissertation, Université Henri Poincaré-Nancy I, 2008.

[12] A. Hartmann, J. M. Lemos, R. S. Costa, J. Xavier, S. Vinga, "Identification of switched ARX models via convex optimization and expectation maximization," Journal of Process Control, vol. 28, pp. 9–16, 2015.

[13] G. Pillonetto, "A new kernel-based approach to hybrid system identification,"Automatica, vol. 70, pp. 21–31, 2016.

[14] A. L. J. Juloski, S. Paoletti, J. Roll, "Recent techniques for the identification of piecewise affine and hybrid systems," in Current trends in nonlinear systems and control, Springer, 2006, pp. 79–99.

[15] S. Paoletti, A. L. J. Juloski, G. Ferrari-Trecate, R. Vidal, "Identification of hybrid systems: A tutorial," European journal of control, vol. 13, no.2, pp. 242–260, 2007.

[16] A. Garulli, S. Paoletti, A. Vicino, "A survey on switched and piece-wise affine system identification," IFAC Proceedings Volumes, vol. 45, no.16, pp. 344–355, 2012.

[17] F. Lauer, G. Bloch, "Switched and piecewise nonlinear hybrid sys-tem identification," in International Workshop on Hybrid Systems: Computation and Control, Berlin., 2008, pp.330–343.

computation in neural systems 6 (3) (1995) 469–505.

[32] M. E. Tipping, Bayesian inference: An introduction to principles and practice in machine learning, in: Advanced lectures on machine Learning, Springer, 2004, pp. 41–62.

[33] T. V. Gestel, J. A. Suykens, G. Lanckriet, A. Lambrechts, B. D. Moor,J. Vandewalle, Bayesian framework for least-squares support vector machine classifiers, Gaussian processes, and kernel Fisher discriminant analysis, Neural computation 14 (5) (2002) 1115–1147.

[34] P. Grünwald, A tutorial introduction to the minimum description length principle, Advances in minimum description length: Theory and applications (2005) 3–81.

[35] F. Lauer, R. Vidal, G. Bloch, A product-of-errors framework for linear hybrid system identification, in: Proc. of the 15th IFAC Symp. on System Identification (SYSID), Saint-Malo, France, 2009.