یازدهمین کنفرانس مهندسی برق ایر
۱۸-۱۶ اردیبهشت ۱۳۸۲

ارائه یک روش جدید برای آموزش بهینهٔ مدل پنهان مارکف

جهانشاه کبودیان *
kabudian@ce.aut.ac.ir

محمدرضا میبدی +
meybodi@ce.aut.ac.ir

محمدمهدی همایون‌پور *
homayoun@ce.aut.ac.ir

* آزمایشگاه سیستمهای هوشمند صوتی-گفتاری،
+ آزمایشگاه محاسبات نرم،
دانشکدهٔ مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران

چکیده

یکی از ابزارهای بسیار قدرتمند در پردازش فرآیندهای اتفاقی و دنباله‌های تصادفی مدل پنهان مارکف یا HMM میباشد. مشهورترین روش آموزش مدل پنهان مارکف روش بام-ولش BW[1] است که یک روش آموزش(جستجوی) محلی بوده و در دام بهینه‌های محلی گرفتار می‌آید. در این تحقیق از روشهای جستجوی سراسری مبتنی بر سرد کردن فلزات (SA[2]) استفاده کردیم. همچنین یک روش SA حافظه دار و کاملاً جدید بنام MiPSA[3] نیز ارائه داده‌ایم که دارای راندمان بالاتری نسبت به SA های بدون حافظه است. آزمایشها نشان داد که کارآیی متوسط الگوریتم BW بیش از کارایی متوسط روشهای سراسری در تعداد تکرارهای نه خیلی زیاد (۱۰۰۰۰ تکرار) است و این به دلیل قدرت بالای روش BW در تنظیم دقیق پارامترها، سرعت آن و پایه ریاضی مستحکم آن می‌باشد . برای رسیدن به بهینه سراسری در تعداد تکرارهای نه خیلی زیاد از روشهای ترکیبی* استفاده کردیم که هم از روش جستجوی سراسری مبتنی بر SA و هم از روش جستجوی محلی BW بهره می‌گیرد و دارای راندمان بالاتری نسبت به روش BW به تنهایی است. به نظر می‌رسد که با استفاده از روش بهینه‌سازی و جستجوی ترکیبی توانسته‌ایم به بهینه سراسری و آموزش بهینه مدل پنهان مارکف نزدیک شویم.

۱- مقدمه

مدل پنهان مارکف یکی از ابزارهای بسیار قوی در پردازش فرآیندهای اتفاقی و دنباله‌های تصادفی است. مشهورترین و پرکاربردترین روش آموزش مدل پنهان مارکف روش بام- ولش است که در واقع یک روش جستجوی محلی است و مانند روشهای دیگر جستجوی محلی در دام بهینه‌های محلی گرفتار می‌آید. برای رفع این ایراد، از روشهای جستجو و بهینه‌سازی سراسری برای آموزش بهینه مدل مارکف استفاده شده است که می‌توانند به راحتی از بهینه‌های محلی بگریزند، ولی این روشها در تنظیم دقیق پارامترهای مدل مارکف و داشتن دقت بسیار بالا، ضعیف هستند و کُند می‌باشند. به همین دلیل روشهای ترکیبی برای جستجوی سراسری پیشنهاد شده‌اند که هم از قابلیت روشهای سراسری و هم از قابلیت روشهای محلی توأماً استفاده می‌کنند. هدف از این کار تحقیقی، مقایسه روشهای محلی، روشهای سراسری و روشهای ترکیبی برای آموزش مدل پنهان مارکف می‌باشد. همچنین یک روش جدید نیز برای جستجو و بهینه‌سازی سراسری ارائه شده است.

در بخش دوم از این مقاله روش BW برای آموزش مدل پنهان مارکف را توضیح خواهیم داد. در بخش سوم به معرفی روشهای بهینه‌سازی سراسری و روش جدیدی که ارائه داده‌ایم، می‌پردازیم. در بخش چهارم، روشهای ترکیبی را بررسی می‌کنیم و در بخش پنجم، به ارائه و بررسی نتایج آزمایشها خواهیم پرداخت.

۲- آموزش مدل پنهان مارکف با الگوریتم BW

در بکارگیری مدل پنهان مارکف، فرض می‌کنیم یک یا چند نمونه از دنباله‌های مشاهدات توسط یک مدل ناشناخته λ تولید شده است و هدف پیدا کردن پارامترهای مدل λ یعنی پارامترهای $b(\ ), A, \pi$ است. مشهورترین و پراستفاده ترین روش برای آموزش مدل پنهان مارکف، الگوریتم BW با

[1] Baum-Welch
[2] Simulated Annealing

تخمین آماری ML[5] است و در واقع فرمولاسیون BW، همان فرمولاسیون الگوریتم بیشینه کردن امید ریاضی یا EM[6] می‌باشد[1]. روش EM با همه قدرت و توانایی و پایه ریاضی مستحکم نسبت به خیلی از روشهای آموزش دیگر، دارای ضعف‌هایی است که باعث شده نسخه‌های پیشرفته‌ترِ آن نیز در سالهای اخیر پیشنهاد شود[2,3].

روش بام- ولش به این صورت عمل می‌کند که ابتداتاً یک حدس اولیهٔ λ_0 بصورت اتفاقی و یا با استفاده از روشهای ابتکاری برای مدل تولید می‌شود. آنگاه در هر مرحله، یک تخمین از پارامترهای مدل در مرحله قبل و با توجه به رشته مشاهدات بدست می‌آید و این تکرار و تخمین آنقدر ادامه می‌یابد تا مدل همگرا شود.

الگوریتم BW نیز مانند روشهایی مثل حرکت در عکس راستای گرادیان GD[7] (انتشار خطا به عقب در شبکه‌های عصبی MLP) یک روش محلی است، ولی برخلاف روشهایی مثل GD تضمین می‌کند که در هر تکرار، راندمان سیستم بهبود می‌یابد و حرکت آن بسیار حساب شده تر و سریعتر است، ولی در GD ممکن است خطای سیستم در یک تکرار نسبت به تکرار قبل بیشتر شود. نکته بسیار مهمی که در مورد الگوریتم BW وجود دارد این است که بر حسب اینکه نقطه شروع اولیه چه باشد، این الگوریتم به بهینه‌های محلی مختلف همگرا خواهد شد و همانطور که در بخش آزمایشها خواهیم دید، ممکن است یک نقطه شروع اولیهٔ B نسبت به نقطه شروع اولیهٔ A دارای $P(O|\lambda)$ کمتری باشد(بدتر باشد)، ولی الگوریتم BW با شروع از نقطه اولیهٔ B به بهینهٔ محلیِ نهایی بهتری نسبت به شروع از نقطه اولیهٔ A برسد. در نظر گرفتن این نکته تأثیر بسیار زیادی در انتخاب نوع الگوریتم خواهد داشت. برای فرار از به دام افتادن در بهینه‌های محلی باید از روشهای جستجوی سراسری یا بهینه سازی سراسری استفاده کرد. مرجع [1] جزئیات بیشتری از الگوریتم BW را ارائه می‌نماید.

۳- روشهای جستجو و بهینه‌سازی سراسری

برای حل مسائل بصورت نزدیک به بهینهٔ سراسری[8] از روشهای جستجوی سراسری[9] یا بهینه‌سازی سراسری استفاده می‌شود. از جمله این روشها میتوان به روش سرد کردن فلزات(SA)، روش الگوریتمهای ژنتیکی(GA)[4] ، روش استراتژی تکاملی(ES)[4] و روش برنامه‌ریزی تکاملی(EP)[4] اشاره کرد. برای بهینه‌سازی سراسری روشهای دیگری نیز وجود دارد[5]. به تازگی نیز یک روش اتوماتون یادگیر تقویتی با مقدار عمل پیوسته(CARLA) برای بهینه‌سازی سراسری در محیطهای پیوسته ابداع شده است[6,7]. در برخی از این روشها، مثل روش SA رسیدن به بهینه سراسری از دید آماری تضمین شده است، ولی در برخی دیگر همچون GA ، ES و EP رسیدن به بهینه سراسری تضمین نشده است. در روش CARLA نیز رسیدن به بهینه سراسری تضمین نمی‌گردد. در تحقیقات جامعی که در موسسهٔ CalTech انجام شده است[8,9]، الگوریتمهای ژنتیکی(GA) با روشهای پیشرفتهٔ مبتنی بر سرد کردن فلزات(SA) مقایسه شده است و در اکثر موارد الگوریتمهای مبتنی بر SA نسبت به GA جوابهای بهتری را در مدت زمان کمتری ارائه کرده‌اند و در بقیه موارد کارآیی این دو یکسان بوده است(SA همیشه سریعتر از GA بوده است). این آزمایشات بازای توابع محدب یا مقعر، توابع پیوسته یا گسسته، توابع خطی یا غیرخطی، توابع با تعداد پارامترهای زیاد یا کم، توابع با تعداد بهینه‌های محلی کم یا بسیار زیاد و حتی توابعِ قطعی یا تصادفی(توابع هزینه همراه با نویز) انجام شده است. مواردی که جواب GA و SA یکسان بوده است، به ازای توابع هزینه‌ای بوده که دارای پارامترهای کم(۲ تا ۳ پارامتر) بوده و توابع مذکور پیوسته و دارای تعداد مینیمم‌های محلی کم بوده‌اند. نکته بسیار مهمی که در این آزمایشها مشاهده شد، این است که تفاوت کارآیی GA در دفعات مختلف اجراء(PerformanceVariance) بیشتر از SA بوده است. البته در آزمایشهای مذکور از نسخه‌های پیشرفته و سریعِ SA استفاده شده است.

از طرف دیگر روشی مثل GA مخصوصاً برای مسائل با پارامترهای پیوسته دارای دقت و رزولوشن پایینی است. این پدیده می‌تواند به این دلیل باشد که در GA مقادیر حقیقی با رشتهٔ بیتها کد می‌گردند و احتمال جهش برای بیت‌های کم‌ارزش و بیت‌های پرارزش یکسان است. یعنی در واقع احتمال تغییر بسیار زیاد یا تغییرِ بسیار کمِ مقادیر پارامترها یکسان است. در حالیکه در روشهایی مثل SA چنین نیست و با رسیدن به جوابهای بهتر و کاهش دما، رزولوشن نیز افزایش می‌یابد و احتمال تغییرات بسیار کم بیش از تغییرات زیاد است. در استراتژی‌های تکاملی (ES) و در برنامه‌ریزیِ تکاملی(EP) با بهره‌گیری از توزیعهای نرمال پیوسته، این ایراد برطرف شده است، ولی ایراد دیگری که وجود دارد، این است که در ES و EP مقدار رزولوشن یا واریانس توزیع یا همان اندازهٔ گام(Step Size) بطور اتوماتیک تطبیق می‌یابد و سازگاری پیدا می‌کند و این سازگاری ممکن است بسیار کند باشد.

یکی از ایراداتی که به SA وارد است این است که از قابلیت توازی برخوردار نیست و به همین خاطر نسخه‌های موازی از SA پیشنهاد گردیده که این ایراد را برطرف میکند[10]. همچنین SA یک روش حافظه‌دار نمی‌باشد که باز هم روشهای حافظه‌دار مبتنی بر SA ابداع گردیده است[11]. هر چند همانطور که گفته شد نسخه‌های غیر موازی و بدون حافظه از SA بنا بر مشاهدات دارای کارآیی بهتری نسبت به GA بوده‌اند[8,9].

با تکیه بر موارد ذکر شده در بالا در این تحقیق بیشتر بر روی روشهای مبتنی بر SA تمرکز شده و این روشها بهینه گردیدند. یک روش جدید حافظه‌دار مبتنی بر SA نیز پیشنهاد گردیده است.

---

[5] Maximum Likelihood

[6] Expectation Maximization

۳-۱ روش سردکردن فلزات (SA)

روش آب‌دادن فلزات یا روش سردکردن فلزات(SA)، یکی از روشهای جستجو و بهینه‌سازی سراسری است که بر اصول ترمودینامیک مبتنی است و از دید آماری تضمین می‌کند که در صورت رعایت کردن روند کاهش دما(Temperature Schedule) بهینهٔ سراسری را پیدا خواهد کرد[12]. این روش کاربرد بسیار زیادی هم در مسائل گسسته و هم در مسائل پیوسته دارد[8,9]. طرز کار SA به این صورت است که در هر تکرار با استفاده از یک تابع چگالی احتمالِ تولید[10] g(x) ، یک نقطه تصادفی را در حول و حوش نقطه جواب قبلی، تولید می‌کند، تابع هزینه E(x) را بازای این نقطه محاسبه می‌نماید و با استفاده از یک تابع احتمال پذیرش[11] h(x) نقطه جدید را بطور احتمالی قبول یا رد می‌نماید. اگر نقطه جدید دارای راندمان بهتری نسبت به نقطه قبلی باشد، احتمال قبول آن بیشتر است. البته در ابتدای جستجو که مقدار دما زیاد می‌باشد و محدوده جستجو وسیع است، احتمال قبول و رد نقطه جدید یکسان است(1/2). با کاهش دما و نزدیک شدن به نقاط بهینهٔ پیدا شده، محدوده جستجو برای بالا بردن دقت و رزولوشن، کمتر می‌گردد. در بخشهای بعدی، روشSA و چند نسخهٔ پیشرفته آن را ارائه خواهیم کرد.

۳-۱-۱ روش SA استاندارد یا BA

روشSA استاندارد یا BA به این صورت عمل می‌کند[8,9] :

۱- مقدار اولیه برای دما (T_1) و یک نقطهٔ اولیه (x_1) انتخاب می‌شود.

۲- $x^* = x_1 \quad , \quad k = 1$

۳- با استفاده از تابع چگالی احتمال تولید g(x) ، یک نقطهٔ جدید x بطور تصادفی در همسایگی x^* انتخاب می‌شود.

$$g(x,x^*) = (2\pi . T_k)^{\frac{-D}{2}} . \exp\left(-\frac{\| x - x^* \|^2}{2T_k}\right) \qquad (۱)$$

۴- تابع هزینه به ازای نقطه x محاسبه می‌شود E(x) .

۵- نقطه جدید x با احتمال h(x) پذیرفته می‌شود (در صورت پذیرش $x^* = x$ است و در غیر اینصورت x^* تغییر نمی‌کند).

$$\Delta E = E(x) - E(x^*) \qquad (۲)$$

$$h(x) = \frac{1}{1 + \exp\left(\frac{\Delta E}{T_k}\right)} \qquad (۳)$$

۶- $k = k + 1$

۷- دما کاهش می‌یابد. $T_k = \frac{T_1}{\ln k}$

۸- اگر شرط خاتمه برقرار است، الگوریتم خاتمه می‌یابد وگرنه قدم ۳ اجراء می‌شود.

در الگوریتم فوق D تعداد پارامترهای تابع هزینه است و منظور از $\| x - x^* \|$ فاصله اقلیدسی دو بردار D بعدی x , x^* می‌باشد. تابع پذیرش یا h(x) در بیشتر مواقع به شکل فوق است که به تابع Barker معروف است ولی به جای آن از تابع متروپولیس نیز می‌توان استفاده کرد[13]. بعضی اوقات از SA بصورت قطعی نیز استفاده می‌شود، یعنی اگر نقطه جدید نقطه بهتری باشد، حتماً قبول خواهد شد و در غیر اینصورت حتماً رد خواهد شد. در الگوریتم SA استاندارد، تابع چگالی احتمال تولید نقطه جدید از نوع گاوسی یا نرمال است و الگوریتم متناظر با چنین تابع چگالی احتمالی الگوریتم BA نامیده می‌شود. اگر روند کاهش دما (قدم ۷ از الگوریتم) از روند ذکر شده سریعتر نباشد، این روش تضمین می‌کند که از دید آماری حتماً بهینه سراسری را پیدا خواهد کرد. همانطور که دیده می‌شود، روش BA با روند کاهش دمای متناظر با آن (قدم ۷) بسیار کُند است و روند رسیدن به جواب بسیار آهسته می‌باشد و به تکرارهای بسیار زیادی نیاز دارد. به همین دلیل نسخه‌های سریع و بسیار سریع از این الگوریتم پیشنهاد شده‌اند. اگر بخواهیم از روش SA برای بهینه‌سازی سراسری مدل پنهان مارکف استفاده کنیم، می‌توان تابع هزینه را چنین تعریف کرد:

$$E = -Log\ P(O|\lambda) \qquad (۴)$$

۳-۱-۲ روش SA سریع یا FA[12]

همانطور که در قسمت قبلی ذکر شد، روند کاهش دما در الگوریتم SA استاندارد یا BA بسیار کند است و رسیدن به رزولوشن زیاد و دقت زیاد و دماهای کم برای حل مسائل پیوسته بسیار زمان بر خواهد بود. برای حل این مشکل یک الگوریتم سردکردنِ سریع یا FA پیشنهاد شده است[8,9] که برای تولید نقاط تصادفیِ جدید در حول و حوش جواب قبلی از توزیع کوشی استفاده می‌کند. به همین خاطر این روش را روش CA[13] نیز می‌نامند. یک تابع چگالی احتمال در فضای D بعدی از نوع کوشی برای تولید بردارهای تصادفی D بعدی بدین شکل می‌باشد[14] :

$$g(x,x^*)=\frac{\Gamma\left(\frac{D+1}{2}\right)}{\pi^{\left(\frac{D+1}{2}\right)}}\cdot\frac{T}{\left(\|x-x^*\|^2+T^2\right)^{\left(\frac{D+1}{2}\right)}} \qquad (۵)$$

برای تضمین شرط رسیدن به نقطه بهینه سراسری لازم است که روند کاهش دما از روند زیر سریعتر نباشد:

$$T(k)=\frac{T_0}{k} \qquad (۶)$$

در روش FA روند کاهش دما از روند فوق نیز می‌تواند سریعتر و بصورت زیر باشد[15]:

$$T(k)=\frac{T_0}{k^{\alpha}} \qquad 1\le\alpha<2 \qquad (۷)$$

همانطور که ملاحظه می‌گردد، روند کاهش دما در FA بسیار سریعتر از BA می باشد. در عمل چون الگوریتم سریعی برای تولید یک بردار تصادفی D بعدی با توزیع کوشی وجود ندارد و نیز به این دلیل که نیاز داریم دماهای بعدهای مختلف از پارامترها متفاوت باشند، لذا در هر بعد از فضای D بعدی بطور مستقل از یک توزیع کوشی یک بعدی استفاده می‌شود که بصورت زیر است:

$$g_i(x_i,x_i^*)=\frac{1}{\pi}\cdot\frac{T_{ik}}{\left(x_i-x_i^*\right)^2+T_{ik}^2} \qquad (۸)$$

برای تضمین شرط رسیدن به بهینه سراسری، باید روند زیر برای کاهش دما رعایت گردد.

$$T_i(k)=\frac{T_{i0}}{k^{\frac{1}{D}}} \qquad (۹)$$

۳-۱-۳ روش SA بسیار سریع یا VFA[14]

در این روش[8,9] باز هم سعی خواهد شد که روند کاهش دما نسبت به روشهای قبلی تسریع گردد. اگر فرض کنیم که x_{ik}^* (مقدار قبلی پارامتر بعد i ام در تکرار k ام) در فاصله $[A_i,B_i]$ قرار گیرد، مقدار جدید این پارامتر را در تکرار k+1 ام اینچنین تولید می‌کنیم:

$$x_{i(k+1)}=x_{ik}^*+y_i(B_i-A_i) \qquad (۱۰)$$

متغیر y_i یک متغیر تصادفی است که محدود و در فاصله $[-1,+1]$ می‌باشد و با تابع چگالی احتمال زیر تولید می‌شود:

$$g_i(y_i)=\frac{1}{2\left(|y_i|+T_{ik}\right)\ln\left(1+\frac{1}{T_{ik}}\right)} \qquad (۱۱)$$

برای تضمین شرط رسیدن به بهینه سراسری، باید روند کاهش دما از روند زیر سریعتر نباشد:

$$T_i(k)=\frac{T_{i0}}{\exp\left(c_i k^{\frac{1}{D}}\right)} \qquad (۱۲)$$

پارامتر c_i که یک عدد مثبت است، برای کارکردِ بهینه سیستم تنظیم می‌شود. همانطور که ملاحظه می‌گردد، روند کاهش دما در VFA بسیار سریعتر از FA ، BA است و رسیدن به دقت و رزولوشن بالا در تعداد تکرار کمتری نسبت به روشهای قبلی میسر می‌گردد(البته با تضمین شرط رسیدن به بهینه سراسری).

۳-۲ یک روش SA جدید حافظه‌دار(MiPSA)

یکی از معایب روش SA و انواع مختلف آن که در بخش‌های قبلی معرفی گردید، این است که روش SA حافظه‌دار نمی‌باشد و این روش فقط ۱ نقطه از جوابهای قبلی را که بصورت احتمالی، بهترین نقطه می‌باشد، در حافظهٔ خود ذخیره می‌کند. البته روشهایی مبتنی بر SA و از نوع حافظه پیشنهاد شده‌اند[11]. در اینجا یک روش SA حافظه‌دار و جدید معرفی می‌گردد که در عین سادگی دارای کارایی بالاتری نسبت به روشهای قب است. در این روش جدید، ما از یک تابع چگالی احتمال مخلوط(Mixture PDF) برای تولید نقاط جدید استفاده می‌کنیم و به همین دلیل این رو را MiPSA (Mixture Pdf SA) نامیده‌ایم. این روش بصورت احتمالی M نقطه از بهترین جوابهای قبلی را در حافظهٔ خود ذخیره می‌کند و تا چگالی احتمالِ مخلوط g(x) برای تولید نقطه جدید x به این صورت می‌باشد:

$$g(x) = \sum_{m=1}^{M} w_m g(x, x^{*(m)}) \qquad (۱۳)$$

در رابطه فوق $x^{*(m)}$ نقطه m ام از مجموعهٔ M عضوی بهترین جوابهای قبلی می‌باشد و $g\left(x, x^{*(m)}\right)$ یک تابع چگالی احتمال با برد میانگین $x^{*(m)}$ است. این تابع چگالی احتمال می‌تواند هرکدام از توابع تولید بکار رفته در روشهای VFA , FA , BA باشد، ولی نکته قابل توجه ای است که بازای هر کدام از انتخاب‌ها، باید روند متناظر آن را برای کاهش دما انتخاب کنیم. بعنوان مثال اگر تابع چگالی احتمال کوشی انتخاب شود، رو کاهش دما نیز باید روند متناظر با همان روش(FA) باشد. برای این که سطح زیر منحنی تابع چگالی احتمال مخلوط یک گردد، باید داشته باشیم:

$$\sum_{m=1}^{M} w_m = 1 \qquad (۱۴)$$

w_m را بصورت ابتکاری از رابطه زیر بدست می‌آوریم:

$$w_m = \frac{c_m}{\sum_{i=1}^{M} c_i} \qquad (۱۵)$$

$$c_m = 1 + \alpha \frac{E_{\max} - E\left(x^{*(m)}\right)}{E_{\max} - E_{\min}} \qquad (۱۶)$$

$E_{\max}$ مقدار هزینه به ازای بدترین عضو از مجموعه M عضوی S (مجموعهٔ بهترین جوابهای قبلی) است. $E_{\min}$ مقدار هزینه به ازای بهترین عضو از مجموعه S می‌باشد . همانطور که ملاحظه می‌شود، شانس بهترین عضو از این مجموعه $(1+\alpha)$ برابرِ شانس بدترین عضو از این مجموعه می باشد. α یک عدد مثبت و بعنوان مثال برابر ۱ می‌تواند باشد. پس از تولید نقطهٔ جدید، این نقطه را با احتمالِ پذیرشِ h(x) می‌پذیریم. در صورت پذیرش، نقطه x جایگزینِ بدترین عضو از مجموعه M عضوی جوابهای قبلی (S) خواهد شد.

$$h(x) = \frac{1}{1 + \exp\left(\frac{\Delta E}{T_k}\right)} \qquad (۱۷)$$

$$\Delta E = E(x) - E_{\max} \qquad (۱۸)$$

همانطور که ملاحظه می‌شود، در روش MiPSA هر چقدر تابع هزینه بازای یکی از نقاط کمتر باشد، مقدار وزن آن و شانس آن بیشتر است و نیز اینکه اگر نقطهٔ جدید پذیرفته شود، جایگزین بدترین نقطه از جوابهای قبلی خواهد شد. این خواص به اضافه خاصیت‌های بسیار خوب قبلی SA (ازجمله تضمین رسیدن به بهینهٔ سراسری)، روش MiPSA را برای بهینه‌سازی سراسری بسیار مناسب می‌سازد.

۴- روشهای ترکیبی برای بهینه سازی سراسری

یکی از مشکلات روشهای بهینه‌سازی سراسری این است که دقت و رزولوشن این روشها کم است. این گونه روشها در تکرارهای بسیار زیاد می‌توانند بهینه سراسری را پیدا کنند، ولی در تکرارهای محدود و متوسط قادر به این کار نیستند. در واقع در روشهای جستجوی سراسری برداشتن قدم بعدی بصورت تصادفی است و بر اساس یک آینده‌نگری دقیق نمی‌باشد. در عوض روشهای جستجوی محلی قدمهای بعدی را براساس یک توجیه ریاضی و بصورت حساب شده بر می‌دارند. از این دید روشهای جستجوی محلی مثل BW بهتر از روشهای جستجوی محلی مثل حرکت در راستای عکس گرادیان (GD) است، زیرا در روش BW تضمین شده است که برداشتن گام بعدی باعث بهبود سیستم می‌گردد، ولی در روشی مثل GD این تضمین وجود ندارد، مگر اینکه نرخ یادگیری بطور دقیق تنظیم گردد که تنظیم دقیق آن برای تضمین بهبود سیستم کاری بسیار مشکل است. بطور خلاصه در روشهای جستجوی سراسری برداشتن گام بعدی بصورت تصادفی است و در یک راستای حساب شده در فضای چند- بعدی پارامترها نیست. از این رو

روشهای جستجوی سراسری در تنظیم دقیق پارامترها[۱۵] و داشتن دقت و رزولوشن بالا ضعیف هستند، در عوض می‌توانند از دام بهینه‌های محلی به راحتی بگریزند. از نقطه نظر دقت و رزولوشن نیز روشهای جستجوی سراسری متفاوت هستند. بعنوان مثال روشهای مبتنی بر SA همیشه با گذشت زمان دقت و رزولوشن خود را افزایش می‌دهند(واریانس توزیع کاهش می‌یابد)، ولی در روشهای مبتنی بر تکامل مانند ES و EP دقت و رزولوشن سیستم بطور خودکار تنظیم شده و سازگاری پیدا می‌کند که این سازگاری ممکن است زمان زیادی بطول بیانجامد. روش GA هم دارای دقت کمتری نسبت به EP و EP برای تنظیم دقیق پارامترهاست[17]. در روشی مثل CARLA نیز دقت نسبی سیستم همواره یکسان است(برابر g_w است). روشهای جستجوی سراسری در تعداد تکرار بی‌نهایت بهینه سراسری را پیدا خواهند کرد(در صورت تضمین آماری مانند روشهای مبتنی بر SA) ولی در تکرارهای محدود و متوسط قادر به این کار نیستند. از این رو محققین روشهای جستجوی ترکیبی را پیشنهاد کرده‌اند. در این روشها هم از قابلیت فرار از بهینه‌های محلی در روشهای سراسری، و هم از قابلیت تنظیم دقیق پارامترها و سرعت و دقت بالای جستجو در روشهای محلی استفاده می‌گردد. بعنوان مثال روشهای جستجوی سراسری مثل GA با روشهای محلی تلفیق شده‌اند[17,18,19,20] و یا روشهای جستجوی سراسری مبتنی بر SA با روشهای جستجوی محلی توأماً بکار رفته‌اند[21,22]. روشهای ترکیبی برای جستجو و بهینه‌سازی را روشهای دومرحله‌ای(Two-Phase) نیز می‌نامند [23]. دومرحله‌ای بودن این روشها بدین معنی نیست که هر کدام از مراحل جستجوی محلی و جستجوی سراسری بطور مستقل و فقط یکبار انجام می‌گیرد، بلکه به دفعات مختلف و بصورت متناوب می‌توانیم در جستجوی ترکیبی بین انجام جستجوی محلی و جستجوی سراسری سوئیچ کنیم.

روش جستجوی ترکیبی به این ترتیب عمل می‌کند که مطابق با یک روند جستجوی سراسری(مثل SA) یک نقطه جدید را تولید می‌نماید. پس از تولید این نقطه (نقطه x(n))، این نقطه را به عنوان نقطه شروع فرض کرده و یک جستجوی محلی را با شروع از این نقطه انجام می‌دهد. البته برای اینکه سرعت جستجوی ترکیبی بالا باشد، جستجوی محلی تا رسیدن به همگرایی و بطور کامل انجام نخواهد شد، بلکه به اندازه N_{LS} تکرار انجام می‌شود و در واقع یک جستجوی محلی تقریبی است و پس از انجام N_{LS} تکرار از جستجوی محلی، به نقطه $x_f(n)$ خواهیم رسید. آنگاه هزینه در نقطه $x_f(n)$ محاسبه می‌گردد و نقطه x(n) را با استفاده از هزینه در نقطه $x_f(n)$ ارزیابی می‌کنیم و نقطه x(n) را قبول یا رد می‌نماییم. در واقع هدف جستجوی ترکیبی از این نوع، آن است که یک نقطه در صورتی قبول شود که جستجوی محلی تقریبی با شروع اولیه از آن نقطه به جواب خوبی برسد. پس از قبول یا رد نقطه x(n) بصورت احتمالی، مثلاً در روش SA دما کاهش می‌یابد و سپس نقطه بعدی x(n+1) با استفاده از تابع چگالی احتمالِ تولیدِ جدید، تولید می‌گردد و این رویه تا رسیدن به شرط همگرایی ادامه می‌یابد. همانطور که در بخش آزمایشات نیز خواهیم دید، در اکثر مواقع در روشهای جستجوی محلی مانند BW خوب بودن یا نبودن یک نقطه اولیه در همان تکرارهای اولیه معلوم می‌گردد و برای ارزیابی یک نقطه اولیه لزومی ندارد که جستجوی محلی را تا پایان و تا رسیدن به همگرایی انجام دهیم. نکته‌ای که باید به آن توجه داشت این است که تعداد تکرارهای جستجوی محلی نباید از یک حدی کمتر باشد، زیرا در اینصورت تخمین ما و ارزیابی ما در مورد نقطه شروع اولیه x(n) درست نخواهد بود.

۵- آزمایشها

در این بخش به ارائه نتایج بدست آمده در آزمایشهای انجام شده خواهیم پرداخت. هدف از این آزمایشها آموزش مدل پنهان مارکف با معیار بیشترین شباهت (ML) و توسط روشهای محلی، سراسری و ترکیبی می‌باشد. مدل مارکفی که قرار است آموزش داده شود، یک مدل پنهان مارکف با سه حالت و سه تابع گاوسی در هر حالت با ماتریسهای کوواریانس قطری می‌باشد. اگر تعداد پارامترهای سیستم با فرض قطری بودن ماتریس کوواریانس P باشد، داریم:

$$P=N(1+N+M(2D+1)) \qquad (۱۹)$$

که به ازای N=3 , M=3 , D=2 تعداد پارامترهای مدل برابر ۵۷ پارامتر خواهد شد. آموزش این مدل در حالت رشته‌های مشاهدات چندتایی(Multiple Observation) بوده و برای آموزش آن از ۱۰ رشتهٔ مشاهده با طول T=20 استفاده شده است. هدف، بیشینه کردن $P(O|\lambda)$ است که O مجموعه ۱۰ رشتهٔ مشاهده و λ پارامترهای مدل پنهان مارکف می‌باشد. همانطور که می‌دانیم آموزش مدل پنهان مارکف یک مسالهٔ بهینه‌سازی مقید است، زیرا باید شروط زیر برقرار باشد:

$$v_{ii} > 0 \quad \text{و} \quad \sum_{m=1}^{N} c_{im} = 1 \quad \text{و} \quad \sum_{j=1}^{N} a_{ij} = 1 \qquad (۲۰)$$

v_{ii} برابر با عنصر قطریِ ماتریس کوواریانس و متناظر با واریانس در بعد i ام است. برای تبدیل مسأله مقید HMM به یک مسأله نامقید برای بهینه‌سازی از نگاشتهای ابتکاریِ زیر استفاده کرده‌ایم:

$$a_{ij} = \frac{\exp(a'_{ij})}{\sum_{j=1}^{N} \exp(a'_{ij})} \qquad (۲۱)$$

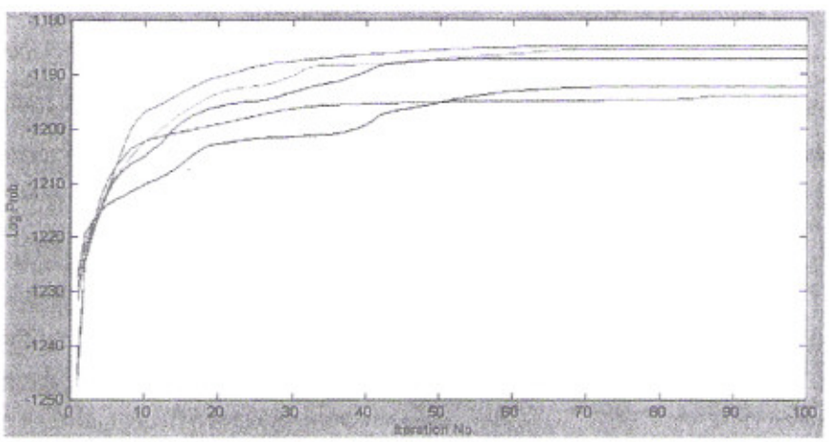
$$_{m} = \frac{\exp(c'_{im})}{\sum_{m=1}^{M}\exp(c'_{im})} \qquad (۲۲)$$

$$_{ii} = \exp(v'_{ii}) \qquad (۲۳)$$

این روابط بدان معنی است که روشهای بهینه‌سازی سراسریِ نامقید همچون SA ، پارامترهای سیستم را در فضای $v'_{ii}, c'_{im}, a'_{ij}, \pi'_{i}$ بصورت نامقید تغییر می‌دهند، ولی تبدیل شدهٔ این پارامترها توسط روابط بالا قیود و شروط مدل مارکف را اقناع میکنند. حال به شرح آزمایشها خواهیم پرداخت.

جدول ۱ - لگاریتم احتمال بازای نقطه شروع و نقطه همگرایی در چهار آزمایش مختلف از روش BW

	آزمایش اول	آزمایش دوم	آزمایش سوم	آزمایش چهارم
نقطه شروع	-۱۲۲۹/۳	-۱۲۳۲/۴	-۱۲۴۵/۲	-۱۲۲۷/۴
نقطه همگرایی	-۱۱۹۳/۴	-۱۱۸۵/۴	-۱۱۸۷/۰	-۱۱۹۲/۲



شکل ۲ - نمودار لگاریتم احتمال برحسب تعداد تکرارها در چهار آزمایش مختلف از انجام روش BW

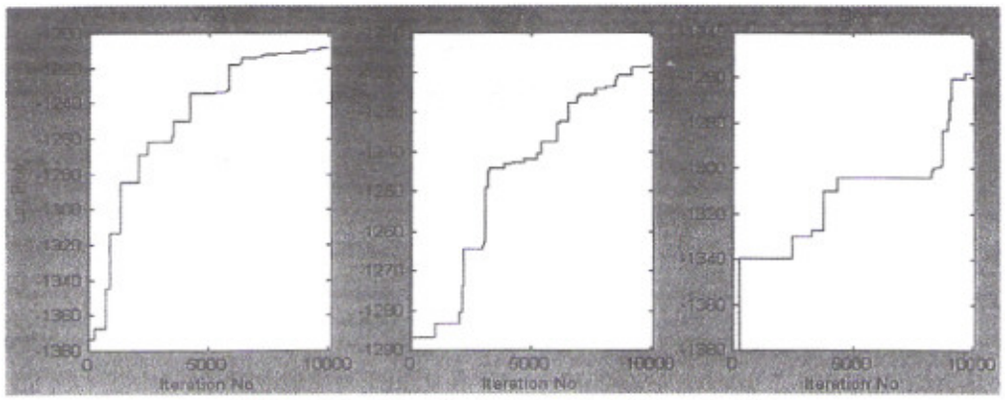
۵-۱ روش آموزش BW

همانطور که قبلاً ذکر شد روش BW یک روش جستجوی محلی است و پر کاربردترین روش آموزش محلی مدلهای HMM می‌باشد. در این قسمت به ازای داده‌های آموزشی مذکور در قسمت قبل، در شکل (۲) از چهار نقطه تصادفی مختلف شروع کرده‌ایم و تا رسیدن به همگرایی تکرارها را ادامه داده‌ایم. مقدار لگاریتم احتمال در نقطه شروع و در نقطه همگرایی در هر کدام از چهار حالت در جدول (۱) ذکر گردیده است. نقطه قابل توجه و مهمی که وجود دارد این است که شروع از یک نقطه اولیهٔ بد ممکن است به یک نقطهٔ همگرایی (نقطه نهایی) خوب منجر شود. این پدیده در مورد روش آموزش بر اساس حرکت در راستای عکس گرادیان(GD) هم در مرجع [17] گزارش شده است. در واقع همین پدیده دید ما را در مورد طراحی الگوریتم‌های جستجو مخصوصاً الگوریتم‌های جستجوی ترکیبی به شدت تحت تأثیر قرار می‌دهد.

همانطور که مشاهده می‌شود با شروع از نقاط شروع مختلف به نقاط همگرایی متفاوتی رسیده‌ایم(داده‌های آموزشی در چهار آزمایش یکسان است) و در واقع به بهینه‌های محلی مختلف همگرا شده‌ایم که این ایراد اصلی روشهای آموزش(جستجوی) محلی است که با استفاده از روشهای جستجوی سراسری و ترکیبی درصدد رفع این ایراد هستیم. نقطه قابل توجه دیگر این است که سرعت روش BW در چند تکرار اول بسیار زیاد است و منحنی لگاریتم احتمال دارای شیب تندی در ابتدای آموزش می‌باشد.

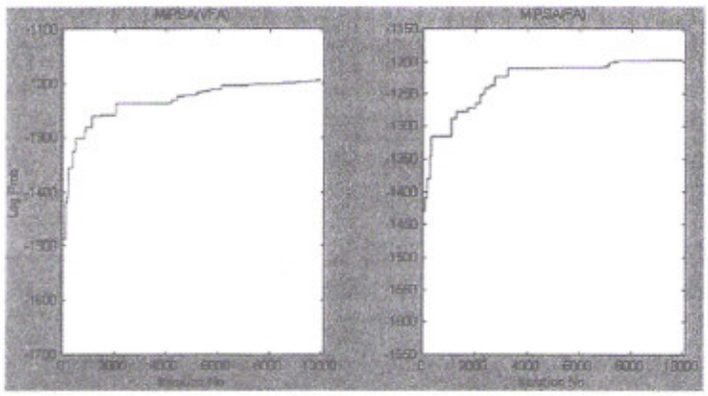
۵-۲ روشهای مبتنی بر سرد کردن فلزات (SA)

از روشهای مبتنی بر SA در این قسمت از سه روش BA ، FA ، VFA استفاده کردیم. در این آزمایشها، قبول یا رد نقطه جدید بطور قطعی بوده است(Deterministic) نه بصورت احتمالی. یعنی اگر هزینه در نقطه جدید کمتر بوده است، نقطه پذیرفته شده و در غیر اینصورت رد می‌شود. مقادیر می‌نیمم و ماکزیمم پارامترها(در روش VFA) و همچنین دماهای اولیه در ابعاد مختلف را با استفاده از اطلاعاتی چون کمترین و بیشترین مقدار پارامترها در بعدهای مختلف، میانگین کل داده‌ها و واریانس کل داده‌ها در بعدهای مختلف بدست آورده‌ایم و مقدار دهی اولیه کرده‌ایم. شکل (۳) نمونه‌ای از عملکرد هر یک از سه روش BA ، FA ، VFA را نشان می‌دهد. همانطور که دیده می‌شود تغییرات لگاریتم احتمال در انتهای جستجو در VFA ، FA کم است و این به آن دلیل است که در این روشها در انتهای جستجو محدوده جستجو بسیار کوچک [illegible]



شکل ۳ - نمونه‌ای از عملکرد سه روش VFA , FA , BA

در شکل (۴) نیز نمونه‌ای از عملکرد روش SA حافظه‌دار(MiPSA) نشان داده شده است. در یکی از اشکال نمونه‌ای از عملکرد روش FA حافظه‌دار و در شکل دیگر عملکرد روش VFA حافظه‌دار ترسیم گردیده است. حافظه سیستم در این دو حالت یعنی M برابر ۱۰ اختیار شده است. یعنی سیستم ۱۰ نقطه از بهترین نقاط بدست آمدهٔ قبلی را در حافظه خود نگه می‌دارد.



شکل ۴ – نمونه‌ای از عملکرد روش FA حافظه‌دار و VFA حافظه‌دار

۳-۵ مقایسه‌ای بین روشهای محلی و سراسری

در این بخش هدف ما مقایسه روشهای محلی و سراسری در بهینه‌سازی و آموزش مدل پنهان مارکف است. در هر کدام از آزمایشها، تعداد تکرارها ۱۰۰۰۰ تکرار بوده است(غیر از BW). برای آنکه نتایج مقایسه یک نتیجهٔ قابل تعمیم و سازگار[۱۶] باشد، هر کدام از روشها به دفعات انجام شده‌اند و شرایط آزمایش برای روشهای مختلف یکسان بوده است. به این منظور برای هر روش، آموزش سیستم ۵۰ بار انجام شد. یعنی ۵۰ نقطهٔ تصادفی برای شروع اولیه و ۵۰ مجموعه آموزشی مختلف، طوری که ۵۰ نقطهٔ شروع اولیه و نیز ۵۰ مجموعهٔ آموزشی مختلف برای تمام روشها یکسان باشد. پس از این مرحله نیز در تولید اعداد تصادفی، بذر اولیه(Seed) برای تمام روشها یکسان قرار داده شد تا شرایط آزمایش کاملاً مساوی باشد. هر آزمایش تا تعداد تکرار برابر با ۱۰۰۰۰ تکرار انجام شد. پس از اتمام آزمایشها، میانگین لگاریتم احتمال بر روی ۵۰ آزمایش را بعنوان میانگین کارآیی روش معیار قرار دادیم. جدول (۲) متوسط راندمان کارآیی(متوسط لگاریتم احتمال) را به ازای روشهای مختلف نشان می‌دهد.

همانطور که ملاحظه می‌گردد، روش محلی BW هنوز هم بیشترین راندمان را داراست. این می‌تواند بدان دلیل باشد که سرعت و دقت روش BW بسیار بالاست و نیز در هر تکرار از الگوریتم BW این تضمین وجود دارد که کارآیی سیستم بهبود خواهد یافت. البته اگر تعداد تکرارها بیش از ۱۰۰۰۰ گردد، روشهای سراسری از یک جایی به بعد مطمئناً از روش BW سبقت خواهند گرفت، ولی فعلاً هدف ما آموزش سیستم در تعداد تکرارهای نه خیلی زیاد می‌باشد. دیده می‌شود که راندمان هر یک از روشهای SA حافظه‌دار(MiPSA) بیش از روشهای SA معمولی نظیر آن می‌باشد و این به دلیل حافظه‌دار بودن روشهای MiPSA می‌باشد. چون روند کاهش دما در BA بسیار کند است و در ۱۰۰۰۰ تکرار نمی‌تواند به دقت و رزولوشن بالایی برسد(به دماهای پایین برسد). به همین دلیل راندمان روش BA پایین‌تر از روشهای FA , VFA است. در روشهای FA ، VFA تابع چگالی احتمال تولید در نقطه فعلی تیز است و دارای دقت بالایی است. در VFA شکل تابع چگالیِ احتمال تولید در نقطه فعلی تیزتر از شکل تابع در روش FA است و دقت آن بیشتر است. اصطلاحاً این نوع از توابع چگالی احتمال برخلاف تابع چگالی احتمال نرمال دارای یک قلهٔ تیز و در عین حال حواشی سنگین یا دُم سنگین[۱۷] هستند. قلهٔ تیز باعث بالا رفتن دقت جستجو می‌شود و حواشی سنگین باعث فرار راحت‌تر از بهینه‌های محلی می‌گردد.

جدول ۲- مقایسه راندمان متوسط روشهای جستجوی محلی و سراسری در ۵۰ آزمایش مختلف

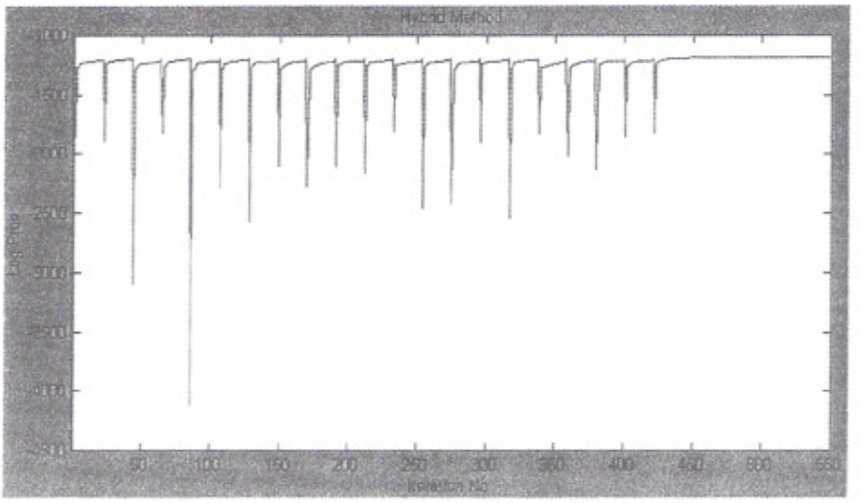
BW	MiPSA(VFA)	MiPSA(FA)	VFA	FA	BA
-۱۱۸۹/۶	-۱۱۹۵/۴	-۱۱۹۸/۵	-۱۲۱۰/۳	-۱۲۱۵/۷	-۱۲۵۹/۱

۶-۵ روش ترکیبی

در این بخش، از الگوریتم جستجوی ترکیبی ذکر شده استفاده خواهیم کرد. روش جستجوی سراسری استفاده شده در این قسمت، روش FA حافظه‌دار یا MiPSA(VFA) است. تعداد تکرارهای جستجوی سراسری ۱۰۰۰ تکرار انتخاب شد و تعداد تکرارهای جستجوی محلی BW در هر تکرار از جستجوی سراسری 15,10,5 تکرار اختیار شد. نتایج آزمایشها در جدول (۳) نشان داده شده است. شکل (۸) نمونه‌ای از عملکرد روش ترکیبی را نشان می‌دهد.

جدول ۳- کارآیی متوسط روش ترکیبی بازای تعداد تکرارهای مختلف در جستجوی محلی

N_{LS} (تعداد تکرارهای BW)	۵	۱۰	۱۵
کارآیی متوسط	-۱۲۰۳/۴	-۱۱۸۱/۳	-۱۱۸۰/۹



شکل ۸- نمونه‌ای از عملکرد روش ترکیبی برای آموزش مدل مارکف

همانطور که ملاحظه می‌شود اگر N_{LS} بیش از ۱۰ تکرار باشد، راندمان جستجوی ترکیبی بیش از راندمان روش BW است. می‌بینیم که اگر N_{LS}، ۵ تکرار باشد، راندمان کمتر از BW است. این بدان دلیل است که با ۵ تکرار از جستجوی محلی BW ، نمی‌توان یک ارزیابی تقریبی درست در مورد نقطه فعلی داشت و تعداد تکرارهای جستجوی محلیِ BW حداقل باید ۱۰ تکرار باشد. اگر در روش جستجوی ترکیبی ۱۰۰۰ تکرار از جستجوی سراسری انجام گیرد و در هر تکرار ۱۰ تکرار از جستجوی محلی انجام شود، آنگاه تعداد تکرارهای کل، برابر همان ۱۰۰۰۰ تکرار در آزمایشهای قبلیِ مربوط به روشهای سراسری خواهد شد(البته یک تکرار از جستجوی محلیِ BW بیش از یک تکرار از جستجوی سراسری زمان می‌گیرد و بیشتر طول می‌کشد). در شکل (۸) نمونه‌ای از عملکرد روش ترکیبی مشاهده می‌گردد. در واقع در این روشِ ترکیبی با کمک روشهای مبتنی بر SA و استفاده از روش محلی BW در یک تعداد تکرارِ نه خیلی زیاد به راندمانی بیشتر از راندمان روشهای محلی BW رسیده‌ایم. راندمان متوسط در حالت $N_{LS} = 15, N_{LS} = 10$ نزدیک به هم است و شاید این بیانگر آن باشد که ما به بهینه سراسری واقعی نزدیک شده‌ایم. قابل توجه آنکه روشهایی مثل SA بسیار سریعتر از روشهایی مثل GA، هستند[8,9]، زیرا در روشهای مبتنی بر SA در هر تکرار یک نقطه (یک راه حل) تولید و ارزیابی می‌گردد، در حالیکه در روشهایی مثل GA ، ES یا EP در هر تکرار، جمعیتی از نقاط(راه‌حل‌ها) باید تولید و ارزیابی شوند.

۶- نتیجه‌گیری

در این مقاله برای آموزش مدل پنهان مارکف، از روشهای جستجوی محلی، جستجوی سراسری و جستجوی ترکیبی استفاده شد. از میان روشهای جستجوی سراسری، روشهای مبتنی بر SA استفاده شدند. آزمایشها نشان دادند که روشهای پیشرفتهٔ مبتنی بر SA بهتر از روش استاندارد SA جواب میدهند. همچنین یک روش جدید حافظه‌دار مبتنی بر SA پیشنهاد و استفاده گردید که نسبت به روشهای SA معمولی(بدون حافظه) دارای کارایی بالاتری است. کارآیی متوسط روش محلی BW از روشهای سراسری فوق در آزمایشهای انجام شده در تعداد تکرارهای نه خیلی زیاد (۱۰۰۰۰ تکرار) بالاتر بوده است و این بدلیل سرعت بالای روش BW دقت [illegible]

محلی BW رسیدیم. روش ترکیبی استفاده شده هم از خاصیت جستجوی وسیع در روشهای سراسری استفاده می‌کند(فرار از بهینه‌های محلی) و هم از سرعت، دقت و رزولوشن بالای جستجوی محلی بهره می‌گیرد.

مراجع

[1] Rabiner, L.R., Juang, B.-H., Fundamentals of Speech Recognition, Prentice-Hall, 1993.

[2] McLachlan, G., Peel, D., Finite Mixture Models, John Wiley & Sons, 2000.

[3] Figueiredo, M.A.T., Jain, A.K., "Unsupervised Learning of Finite Mixture Models", IEEE Trans. on PAMI, Vol. 24, No. 3, pp. 381-896, March 2002.

[4] Bäck, T., Evolutionary Algorithms in Theory and Practice, Oxford University Press, 1996.

[5] Stuckman, B.E., Easom, E.E., "A Comparison of Bayesian/Sampling Global Optimization Techniques", IEEE Trans. on SMC, Vol. 22, No. 5, Sep. 1992.

[6] Howell, M.N., Gordon, T.J., Best, M.C., "The Application of Continuous Action Reinforcement Learning Automata to Adaptive PID Tuning", --- .

[7] Howell, M.N., Gordon, T.J., "Continuous Learning Automata and Adaptive Digital Filter Design", UKACC International Conference on Control, Sep. 1998.

[8] Ingber, L., Rosen, B., "Genetic Algorithms and Very Fast Simulated Reannealing: A Comparison", Mathematical and Computer Modelling, Vol. 16, No. 11, 1992.

[9] Ingber, L., "Simulated Annealing: Practice versus Theory", Mathematical and Computer Modelling, Vol. 18, No. 11, 1993.

[10] Kimura, K., Taki, K., "Time-Homogeneous Parallel Annealing Algorithm", Technical Report, Institute for New Generation Computer Technology, Tokyo, Japan, 1992.

[11] Lo C.-C., Hsu, C.-C., "An Annealing Framework with Learning Memory", IEEE Trans. on SMC, Vol. 28, No. 5, Sep. 1998.

[12] Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., "Optimization by Simulated Annealing", Science, Vol. 220, No. 4598, pp. 671-680, May 1983.

[13] Locatelli, M., "Simulated Annealing Algorithms for Continuous Global Optimization", ***in*** Horst, R., Pardalos, P.M.(Eds.), Handbook of Global Optimization, Kluwer Academic Publishers, Boston, 1995.

[14] Lindsey, J.K., Lindsey, P.J., "Multivariate Distributions with Correlation Matrices for Nonlinear Repeated Measurements", --- .

[15] Fausett, L.V., Fundamentals of Neural Networks, Prentice-Hall, 1994.

[16] Narendra, K.S., Thathatchar, M.A.L., Learning Automata: An Introduction, Prentice-Hall, 1989.

[17] Yao, X., "Evolving Artificial Neural Networks", Proceedings of the IEEE, Vol. 87, No. 9, pp. 1423-1447, Sep. 1999.

[18] Bäck, T., Fogel, D., Michalewics, Z.(Eds.), Handbook of Evolutionary Computation, Institute of Physics Publishing & Oxford University Press, 1997.

[19] Hart, W.E., Adaptive Global Optimization with Local Search, PhD Thesis, University of California at San Diego, 1994.

[20] Renders, J.-M., Flasse, S.P., "Hybrid Methods using Genetic Algorithms for Global Optimization", IEEE Trans. on SMC, Vol. 26, No. 2, Apr. 1996.

[21]-Desai, R., Patil., R., "SALO: Combining Simulated Annealing and Local Optimization for Efficient Global Optimization", Proceedings of the 9th Florida AI Research Symposium (FLAIRS-96), pp. 233-237, 1996.

[22] Lee, Y.-O. et al., "Optical Model Parameter Search with Simulated Annealing and Marquardt-Levenberg Method", Symposium on Nuclear Data, Japan, 2000.

[23] Schoen, F., "Two-Phase Methods for Global Optimization", ***in*** Horst, R., Pardalos, P.M.(Eds.), Handbook of Global Optimization, Kluwer Academic Publishers, Boston, 1995.