آموزش تمایزی مدل مخفی مارکوف با بکارگیری الگوریتم PSO

هدیه ساجدی[×]، حسین ثامتی[†]، حمید بیگی[‡]، باقر باباعلی[§]

چکیده

مدل مخفی مارکوف(HMM) جهت مدلسازی دنباله‌های تصادفی با ساختار حالت متناهی مورد استفاده قرار می‌گیرد. به دلیل وجود الگوریتم‌های قوی و کارای بیشترین درست‌نمایی جهت یافتن پارامترهای HMM، این روش به عنوان یک ابزار مدلسازی موفق مطرح گردیده است. با استفاده از دادگان آموزشی مناسب، پارامترهای یک مدل با ساختار معلوم را می‌توان با تخمین خوبی به دست آورد، اما در عمل تضمینی برای دستیابی به بهینه سراسری وجود ندارد. عدم توجه به دقت بازشناسی یا به عبارت دیگر آموزش مستقل هر مدل، از دیگر معایب این الگوریتم‌های آموزش است. روش‌هایی برای رفع این معایب ارائه شده‌اند، که یا به دلیل مشکلات عددی از کارایی لازم برخوردار نیستند و یا فاقد پایه نظری قوی جهت تضمین همگرایی می‌باشند. در این مقاله الگوریتم PSO جهت آموزش تمایزی مدل مخفی مارکوف پیوسته در یک سیستم بازشناسی گفتار فارسی بکار گرفته می‌شود؛ هدف آن است که با استفاده از قابلیت الگوریتم PSO در جستجوی بهینه سراسری فضای راه حل مساله و با استفاده از مفاهیم آموزش تمایزی، دقت بازشناسی سیستم بهبود یابد. نتایج آزمایشات نشان می‌دهد که با استفاده از این روش، خطای بازشناسی نسبت به آموزش HMM با روش‌های مبتنی بر بیشترین درست‌نمایی، ۳,۳۶ درصد کاهش یافته است.

کلمات کلیدی

مدل مخفی مارکوف ، آموزش تمایزی ، الگوریتم PSO ، بازشناسی گفتار

Discriminative Training of Hidden Markov Model Using PSO Algorithm

Hedieh Sajedi, Hossein Sameti, Hamid Beigy, Bagher Babaali

Abstract

Hidden markov model(HMM) is used for definite structure stochastic sequences modeling. Because of strong and efficient maximum likelihood algorithms for HMM parameters finding, HMM is known as a successful modeling tool. By the known model structure and using the convenient training data set, we can find the model parameters with good estimation. But no guaranty would exist for reaching a global optimum. Another weakness of the classic HMM training algorithms is that Maximum Likelihood measure doesn't concern about recognition accuracy. In other word each model trained separately. Other methods proposed for covering these disadvantages, but because of the high computational complexity have not sufficient performance or have not strong theoretical basis for algorithm convergence guarantying. In this paper PSO algorithm is used for discriminative training of continuous HMM in a persian speech recognition system. The goal is using this algorithm ability to find the global optimum in search space of the problem and improving the recognition preciseness by utilization discrimination concepts. The experimental results showed that by using PSO algorithm, the recognition error rate, comparing to HMM training with the classic maximum likelihood algorithms 3.36 percent reduced.

Keywords

Hidden Markov Model, Discriminative Training, PSO Algorithm, Speech Recognition

[×] دانشجوی دکتری، دانشگاه صنعتی شریف، دانشکده مهندسی کامپیوتر،A_Sajedi@ce.sharif.edu
[†] استادیار، دانشگاه صنعتی شریف، دانشکده مهندسی کامپیوتر،Sameti@sharif.edu
[‡] استادیار، دانشگاه صنعتی شریف، دانشکده مهندسی کامپیوتر،Beigy@sharif.edu
[§] دانشجوی دکتری، دانشگاه صنعتی شریف، دانشکده مهندسی کامپیوتر، Babaali@ce.sharif.edu

۱- مقدمه

مدل مخفی مارکوف به عنوان یکی از موفق‌ترین روش‌ها در بازشناسی گفتار، به ویژه گفتار پیوسته شناخته شده است[3]. الگوریتم‌های قوی و کارایی با معیار بیشترین درست‌نمایی(ML)[1]، جهت یافتن پارامترهای مدل مخفی مارکوف موجود هستند؛ این الگوریتم‌ها با معلوم بودن ساختار مدل[2] و با استفاده از مجموعه‌ی دادگان آموزشی مناسب، می‌توانند پارامترهای مدل را با تخمین خوبی بیابند. در این روش‌ها، جستجو به طور تصادفی از یک نقطه اولیه در فضای مساله آغاز شده و در هر تکرار از الگوریتم، پارامترهای مدل به سوی بهینه محلی سوق داده می‌شوند. حساسیت این الگوریتم‌ها به تنظیمات اولیه‌ی پارامترها، موجب عدم وجود تضمین جهت دستیابی به بهینه‌ی سراسری است. از سوی دیگر، الگوریتم‌های آموزش با معیار بیشترین درست‌نمایی، دربرگیرنده‌ی مفهوم تمایز، که در تشخیص الگو و بالاخص در یک سیستم بازشناسی گفتار مورد توجه می‌باشد، نیستند. بطور کلی در این الگوریتم‌ها، میزان مطلوبیت نقاط همگرایی که اغلب به سرعت و در چند تکرار الگوریتم حاصل می‌شوند، مشخص نیست. لذا الگوریتم‌های دیگری مانند آموزش با معیار بیشترین اطلاعات متقابل[3] و یا روش‌های آموزش تصحیح کننده[4] پیشنهاد شده‌اند که به دلیل مشکلات عددی و یا فقدان اساس نظری قوی جهت تضمین همگرایی الگوریتم، از کارایی بالایی برخوردار نیستند[4] .

PSO[5] یک روش بهینه‌سازی است که از رفتار اجتماعی موجود در طبیعت الهام گرفته شده است. سادگی پیاده‌سازی، مقیاس پذیری در ابعاد و کارایی خوب تجربی از خواص مطلوب این الگوریتم هستند؛ لذا الگوریتم PSO در برنامه‌ریزی غیرخطی و بهینه‌سازی توابع با دامنه‌های پیوسته، گزینه‌ای قابل توجه است.

در این مقاله، با استفاده از الگوریتم PSO که قابلیت جستجوی بهینه سراسری را دارا است و نیز با ایده گرفتن از روش‌های آموزش تمایزی، پارامترهای مدل مخفی مارکوف پیوسته تخمین زده می‌شود؛ به عبارت دیگر با بهینه‌سازی پارامترهای مدل، توصیف دقیق‌تری از دنباله‌ی مشاهدات ارائه می‌گردد.

الگوریتم PSO شامل گروهی[6] از اجزاء[7] است که در فضای جستجوی راه‌حل‌های ممکن مساله، حرکت می‌کنند. در پیاده‌سازی الگوریتم PSO، برای دستیابی به هر مدل متناظر با یک واحد آوایی، یک گروه در نظر گرفته می‌شود و هر جزء گروه نشانگر یک HMM است. روش طراحی گروه‌ها و آموزش مدل‌ها به گونه‌ای است که در هر تکرار الگوریتم، مدل‌ها به نمونه‌های آموزشی متناظر نزدیک و از نمونه‌های آموزشی نامتناظر دور می‌شوند. جهت ارزیابی کیفیت اجزاء گروه (مدل‌های HMM) در بازشناسی واحدهای آوایی، معیارهای سنجش مختلفی در نظر گرفته می‌شود. از جمله ملاک‌های مورد استفاده می‌توان به تعداد نمونه‌های درست بازشناسی شده نسبت به تعداد کل نمونه‌ها، امتیاز به دست آمده با استفاده از معیارهای MMD، MMI و MCE اشاره کرد. با استفاده از این ملاک‌ها به عنوان تابع ارزشیابی، می‌توان از خاصیت جستجوی بهینه سراسری PSO به همراه خاصیت تمایزی معیارها سود جست.

آزمایشات گوناگونی جهت تعیین درصد خطای بازشناسی مدل‌هایی که با استفاده از الگوریتم PSO، بصورت تمایزی آموزش دیده‌اند، انجام پذیرفته و نتایج آن گزارش شده است.

در بخش دوم، آموزش مدل مخفی مارکوف و در بخش سوم الگوریتم PSO مورد بررسی قرار خواهد گرفت. در بخش چهارم چگونگی آموزش HMM با بکارگیری الگوریتم PSO شرح داده می‌شود. بخش پنجم به بررسی آزمایشات انجام شده و نتایج آنها می‌پردازد و در فصل ششم جمع‌بندی و راه‌کارهای آتی ارائه شده است.

۲- آموزش مدل مخفی مارکوف

هدف از آموزش مدل، یافتن پارامترهایی است که مشخصات زمانی و طیفی موجود در یک سیگنال ورودی را تشریح کنند. در مرحله آموزش، تلاش بر آن است که با استفاده از داده‌های آموزشی مختلف برای هر واحد بازشناسی، مدلی ساخته شود. این امر را می‌توان با ساختن مدلی که چگالی احتمال $p(O \mid \lambda)$ را با داشتن دنباله مشاهدات بیشینه می‌کند، تحقق بخشید. از آنجا که فرآیند آموزش داده محور است پایگاه داده نمونه‌ها باید به اندازه کافی غنی باشد. هر چه تعداد و تنوع داده‌های آموزشی مورد استفاده افزایش یابد، امید بیشتری برای تشکیل مدل‌های دقیقتر وجود دارد. به عبارت دیگر، استفاده از داده‌های آموزشی کافی برای تخمین پارامترها، تضمین‌کننده رفتار مدل‌ها و مدلسازی صحیح است.

پس از مقداردهی اولیه مدل مخفی مارکوف، فرآیند دو مرحله‌ای تنظیم زمانی مشاهدات O با حالت[8]های مدل و تخمین مجدد پارامترهای مدل بر اساس توزیع داده‌ای روی حالت‌های مدل آنقدر تکرار می‌شود تا شرط توقف برآورده شود یا به عبارت دیگر آموزش به یک نقطه بهینه محلی همگرا شود. هدف تخمین مجدد در فرآیند آموزش، تخمین دوباره مدل λ است به گونه‌ای که نتیجه حاصل مدلی محتمل‌تر و مناسب‌تر مانند $\bar{\lambda}$ باشد. می‌توان این رویه را به صورت رابطه (۱) تعریف کرد:

$$f : O, \lambda \rightarrow \bar{\lambda} \quad (۱)$$

که درآن باید داشته باشیم : $\bar{P} = p(O \mid \bar{\lambda}) \geq p(O \mid \lambda) = P$ و $p(O \mid \bar{\lambda})$ بیشینه باشد.

متداولترین روش‌های تخمین پارامترهای HMM، بر اساس معیار بیشترین درست‌نمایی عمل می‌کنند. این کاربرد به علت سادگی و قابلیت محاسباتی بالای معیار ML است؛ اما به طور کلی این معیار، در مسائلی که ذات داده‌هایی که باید بازشناسی شوند متفاوت از فرم تابع چگالی بکارگرفته شده در مدل باشد، آسیب پذیر بوده و راه حل بهینه محلی را نتیجه می‌دهد. به طور کلی روش‌های کلاسیک بهینه‌سازی تابع هدف از یک رویه تکراری بر اساس روش جستجوی گرادیان

استفاده می‌نمایند. روش جستجوی گرادیان معمولاً یک بهینه محلی را به عنوان جواب مساله نتیجه می‌دهد. از سوی دیگر در روش‌های کلاسیک آموزش HMM، هر مدل بطور جداگانه و بدون ارتباط با مدل‌های دیگر بگونه‌ای آموزش می‌بیند که تا حد امکان بر نمونه‌های متناظر از مجموعه آموزشی منطبق گردد. اگر چه از دیدگاه تئوریک روش‌های آموزش کلاسیک به مدل‌های با بیشترین دقت بازشناسی همگرا می‌شوند، اما در عمل این مهم تقریباً هیچگاه دست‌یافتنی نیست. دو دلیل عمده، کافی نبودن تعداد نمونه‌های مجموعه آموزشی و پیچیدگی تابع شباهت (خصوصاً توابعی با تعداد زیاد ماکزیمم‌های محلی) است. به بیان دیگر با داده‌های آموزشی محدود، فرآیند آموزش به جذب در بستر یکی از ماکزیمم‌های محلی تابع شباهت منجر می‌شود.

الگوریتم‌هایی مانند حداکثر اطلاعات مشترکMMI [5]، حداقل اطلاعات تمایزی[۹](MDI) [6,7] ، حداکثر فاصله مدلها[۱۰] (MMD) [8] و روش حداقل نرخ خطای طبقه‌بندی[۱۱] (MCE) [9] جهت پوشش ضعف تمایزی نبودن روش‌های مبتنی بر ML، ارائه شده‌اند.

روش‌های ترکیبی متنوعی از الگوریتم‌های آموزش تمایزی مدل مخفی مارکوف و الگوریتم ژنتیکی ارائه شده است. در [۱و۲] روش آموزش ترکیبی مدل مخفی مارکوف با استفاده ازالگوریتم ژنتیکی و معیارهای سنجش تمایزی MMD,MCE و MMI ارائه شده است. نتایج تجربی این تحقیقات نشان داده است که الگوریتم‌های ترکیبی کارایی بهتری نسبت به الگوریتم‌های مبتنی بر ML داشته و بهبود ۸۳ , ۴ درصدی را در یک سیستم بازشناسی گفتار ایجاد نموده‌اند.

۳- الگوریتم PSO

گاهی دسته سیستم‌های پیچیده، سیستم‌های گروهی[۱۲] نامیده می‌شود و شامل روش‌های محاسباتی جدیدی است که می‌تواند برای حل کارا و قابل اعتماد مسائل دشوار بکار گرفته شود. وقتی گروهی در طبیعت، مساله‌ای را حل می‌کنند، قابلیت آنها معمولاً با صفت هوش گروهی شناخته می‌شود. از میان مثال‌های شناخته شده‌ی حشرات اجتماعی، می‌توان به زنبوران عسل و اجتماع مورچگان اشاره نمود. در سال‌های اخیر اثبات شده است که اصول محاسباتی را می‌توان از هوش گروهی، شناسایی، خلاصه و استخراج نمود و در اهداف علمی و صنعتی مورد استفاده قرار داد. یکی از موثرترین الگوریتم‌های سیستم‌های گروهی، الگوریتم PSO است[10]. PSO یک روش مبتنی بر رفتار اجتماعی پرندگان است که در سال ۱۹۹۵ میلادی توسط Kennedy James(روانشناس اجتماعی) وRuss Eberhart (مهندس کامپیوتر) جهت بهینه‌سازی توابع غیرخطی پیوسته ارائه شده است[11]. PSO از یک سو به حیات مصنوعی[۱۳] خصوصاً تئوری‌های گروهی و از سوی دیگر به الگوریتم‌های پردازش تکاملی و به طور خاص به استراتژی تکاملی و الگوریتم ژنتیکی مرتبط است.

الگوریتم PSO شامل گروهی از اجزاء است که در یک فضای جستجوی چند بعدی از راه‌حل‌های ممکن مساله، حرکت می‌کنند. در این فضا یک معیار ارزیابی تعریف می‌شود و سنجش کیفیت راه‌حل‌های مساله از طریق آن صورت می‌پذیرد. در [4,12] استفاده از PSO به جای روش یادگیری پس‌انتشار[۱۴] در شبکه‌های عصبی مورد بررسی قرار گرفته و در اکثر موارد PSO به عنوان روشی سریع و قابل قبول گزارش شده است. همچنین ثابت شده است که PSO برای حل بسیاری از مسائل بهینه‌سازی سراسری کارا است و در برخی موارد به مشکلاتی که بر دیگر تکنیک‌های تکاملی وارد است، دچار نمی‌شود[13]. الگوریتم PSO به سادگی قابل پیاده‌سازی است و از دیدگاه محاسباتی (سرعتCPU و حافظه) کم هزینه است[14]، به علاوه به اطلاعات گرادیانی تابع هدف مفروض نیازی ندارد.

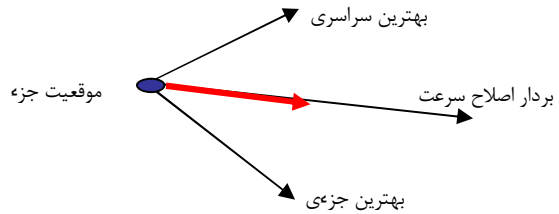
اجزاء گروه در PSO به طور تصادفی مقداردهی اولیه می‌شوند و سپس جستجو برای پیدا کردن جواب بهینه آغاز می‌گردد. در هر تکرار الگوریتم، هر جزء موقعیت بعدی خود در فضای جستجو را با توجه به دو مقدار تغییر می‌دهد؛ یکی بهترین موقعیتی است که خود جزء تاکنون داشته است (pbest) و دیگری بهترین موقعیتی که تاکنون توسط کل اجزاء گروه به دست آمده و در واقع بهترین pbest در کل گروه است(gbest). pbest برای هر جزء در واقع حافظهٔ اتوبیولوژیکی آن محسوب می‌شود. gbest همان دانش عمومی گروه است و وقتی که اجزاء موقعیت خود را بر اساس gbest تغییر می‌دهند، در واقع تلاش می‌کنند که سطح دانش خود را به سطح دانش گروه برسانند. از نظر مفهومی، بهترین جزء گروه همه اعضای گروه را به یکدیگر مرتبط می‌نماید و بهترین موقعیت محلی، توانایی جزء در به خاطرسپاری موقعیت گذشته خود است. هر تکرار از الگوریتم وقتی اتفاق می‌افتد که تمامی اجزاء در تکرار قبل، در فضای جستجو حرکت کرده باشند. با توجه به مقادیر pbest و gbest، هر جزء از روابط (۲) و (۳) برای تعیین موقعیت بعدی خود استفاده می‌کند:

$$v_{t+1} = w_t \cdot v_t + c_1 \cdot \mathrm{rand}()(\mathrm{pbest} - \mathrm{present}_t) + c_2 \cdot \mathrm{rand}()(\mathrm{gbest} - \mathrm{present}_t) \quad (۲)$$

$$\mathrm{present}_{t+1} = \mathrm{present}_t + v_{t+1} \quad (۳)$$

$\mathrm{present}_t$ موقعیت و v_t سرعت حرکت کنونی یک جزء است. در رابطه (۲)، c_1 و c_2 پارامترهای یادگیری هستند که معمولاً برابر با ۲ در نظر گرفته می‌شوند. تنظیم پارامترهای سرعت تا اندازه زیادی کارایی PSO را تعیین می‌کند[15]. در[16] نشان داده شده است که مقادیر تصادفی با میانگین یک، شانس ۵۰ درصدی تغییر بالا و پایین را حول بهترین نقاط pbest و gbest فراهم می‌کند. این تنظیمات علاوه بر تغییرات تصادفی، قابلیت انعطاف جستجو را حفظ می‌نماید. rand() تابع تولید اعداد تصادفی در محدودهٔ [0,1] است. w_t یک

پارامتر کنترلی است که تأثیر سرعت فعلی اجزاء را بر سرعت بعدی آنها کنترل می‌کند و بین توانایی الگوریتم در جستجو بـه صـورت محلـی و جستجو به صورت سراسری تعادل ایجاد می‌نماید. بنابراین منطقـی بـه نظر می‌رسد که w_t به صورت تدریجی کاهش یابد تا در ابتـدا جـستجو به صورت سراسری صورت پذیرد و به تدریج به سمت جستجوی محلی پیش رود[16]. الگوی حرکت هر جزء در گروه بصورت شکل (۱) است:



شکل(۱): الگوی حرکت یک جزء در گروه

بردار موقعیت هر جزء و بالطبع بردارهای pbest، gbest و v_t می‌توانند به صورت چند مؤلفه‌ای باشند که تعداد مؤلفه‌های آنها با توجه به مساله بهینه‌سازی مورد نظر، مشخص می‌گردد. الگوریتم PSO به صورت شکل (۲) خلاصه می‌شود:

```
For each particle
    Initialize particle
End
Do
    For each particle
        Calculate fitness value
        If the fitness value is better than the best fitness value
          (pbest) in history
            set current value as the new pbest
    End
    Choose the particle with the best fitness value of all the
    particles as the gbest
    For each particle
        Calculate particle velocity according equation (1)
        Update particle position according equation (2)
    End
While maximum iterations or minimum error criteria is not
attained
```

شکل(۲): الگوریتم PSO

دشواری تنظـیم پارامترهـای PSO جهـت دسـتیابی بـه کـارایی مطلوب از معایـب ایـن الگـوریتم اسـت[17]. در صـورت عـدم تنظـیم مناسب پارامترها، PSO به سوی بهینـه محلـی متمایـل شـده و دچـار همگرایی زودرس می‌شود[18].

۴- آموزش HMM با استفاده از الگوریتم PSO

در روش ارائه شده متناظر با هر واحد آوایـی، یـک مـدل مخفـی مارکوف چپ به راست N حالته با مشاهدات پیوسته تشکیل مـی‌شـود. هر حالت بصورت M گوسی F بعدی(به تعداد ویژگی‌هـای نمونـه‌هـای آموزشی) در نظر گرفته شده است. به دلیل کمبود نمونه‌های آموزشـی و نیز کاستن حجم محاسبات، مـاتریس کوواریـانس گوسـی‌هـا قطـری فرض شده است. همچنین با توجه به تاثیر ناچیز مقادیر احتمالی پرش بین حالت‌ها در روند آموزش، این احتمالات ثابت فرض شده است.

جهت یافتن پارامترهای هر مدل، یک گروه از اجزاء در نظر گرفته می‌شود و هر جزء در گروه نشان دهنده یک سری مقادیر تخمـین زده شده‌ی پارامترهای HMM است.

بردار موقعیت جزء i ام گروه به صورت رابطـه (۴) در نظـر گرفتـه شده است:

$$X_i = (x_{i1}, x_{i2}, ..., x_{iD}) \qquad (۴)$$

کـه در آن D تعـداد کـل پارامترهـای HMM شـامل میـانگین و واریانس گوسی‌های موجود در هر حالت و اوزان آنهـا اسـت و از رابطـه (۵) محاسبه می‌گردد:

$$D = N \times M \times F \times 2 + N \times M \qquad (۵)$$

نرخ تغییر موقعیت (سرعت) جزء i به صورت رابطه (۶) نشان داده می‌شود:

$$V_i = (v_{i1}, v_{i2}, ..., v_{iD}) \qquad (۶)$$

v_{ik} سرعت بعد k جزء i است.

بهترین موقعیت پیشین (موقعیتی با بهترین میزان کیفیت) جزء i به شکل رابطه (۷) و بهترین موقعیتی که تاکنون توسط گروه به دسـت آمده است، به صورت رابطه (۸) در نظر گرفته شده است:

$$pbest_i = (p_{i1}, p_{i2}, ..., p_{iD}) \qquad (۷)$$

$$gbest_i = (p_{g1}, p_{g2}, ..., p_{gD}) \qquad (۸)$$

موقعیت و سرعت هر جزء توسط بردارهای با اعداد اعـشار شـناور نمایش داده می‌شود. به دلیل آنکه تغییر موقعیت مکانی مولفـه‌هـای از نوع میانگین نـسبت بـه تغییـر موقعیـت مکـانی مولفـه‌هـای از جـنس کوواریانس نقش موثرتری را در عملکرد مدل ایفا می‌کنند، پارامترهای مـوثر در سـرعت مولفـه‌هـای از نـوع میـانگین (c_1,c_2,w_t) متفـاوت از پارامترهـای مولفـه‌هـای کوواریانـسی (cv_1,cv_2,wv_t) در نظـر گرفتـه مـی‌شـود. بـا مولفـه‌هـای از نـوع ضـرایب تلفیـق، مـشابه مولفـه‌هـای کوواریانسی رفتار می‌شود. در هر تکرار از الگوریتم، سرعت هر مولفـه‌ی از جنس میانگین جـزء i مطـابق رابطـه (۹) و سـرعت هـر مولفـه‌ی از جنس کوواریانس و یا ضریب تلفیق با اسـتفاده از رابطـه (۱۰) اصـلاح می‌شود:

$$v_{t+1}(ij) = w_t(ij) \cdot v_t(ij) + c_1 \cdot rand()(pbest(ij) - present_t(ij)) + c_2 \cdot rand()(gbest(ij) - present_t(ij))$$

(۹)

$$v_{t+1}(ij) = wv_t(ij) \cdot v_t(ij) + cv_1 \cdot rand()(pbest(ij) - present_t(ij)) + cv_2 \cdot rand()(gbest(ij) - present_t(ij))$$

(۱۰)

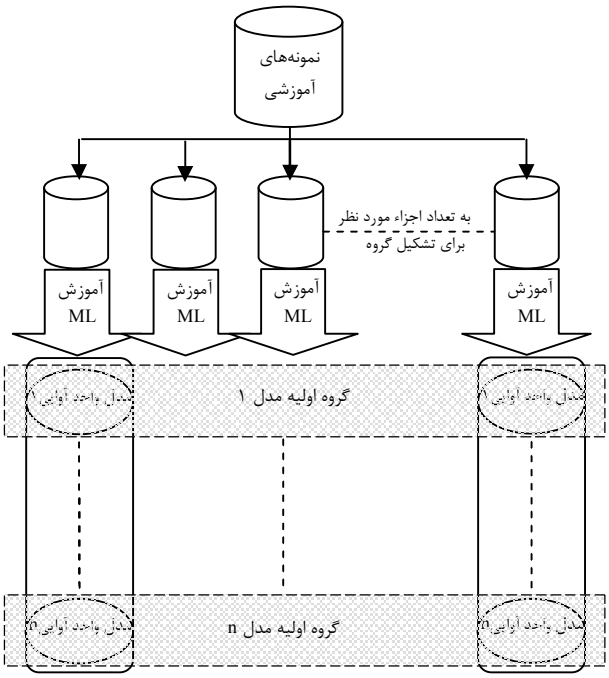
با استفاده از الگوریتم PSO پارامترهای مدل مخفی مارکوف مکرراً اصلاح می‌شوند و احتمال دنباله مشاهدات تا رسیدن به یک نقطه محدود کننده افزایش می‌یابد.

۴-۱- مقداردهی اولیه اجزاء

مقداردهی اولیه پارامترهای مدل، یکی از عوامل تعیین کننده‌ی دقت بازشناسی مدل نهایی است. همچنین تجربه نشان داده است که مقداردهی اولیه مناسب، در مسائل بهینه‌سازی با فضای راه حل پیوسته ضروری است. چنین تخمین اولیه‌ای می‌تواند با استفاده از چندین روش، مانند خوشه‌بندی دستی دنباله‌های مشاهده بر حالتها با متوسط‌گیری مشاهدات داخل حالتها، خوشه بندی با بیشترین درست‌نمایی برای مشاهدات با متوسط‌گیری و خوشه‌بندی به روش k-means حاصل شود.

روش مقداردهی اولیه در این تحقیق در ادامه شرح داده می‌شود. هر واحد در دادگان آموزشی شامل چند فریم است و از هر فریم یک بردار ویژگی استخراج شده است. ابتدا بردار ویژگی از فایل مربوط خوانده می‌شود و با توجه به تعداد حالت‌های مدل متناظر، دنباله‌ی فریم، طوری تقسیم‌بندی می‌شود که به هر حالت، تعداد فریم‌های مساوی منتسب گردد. سپس از فریم‌های اختصاص یافته به هر حالت میانگین‌گیری شده و حاصل، به عنوان یک بردار مشاهده مربوط به آن حالت در نظر گرفته می‌شود. این عمل برای تمامی واحدهای دادگان آموزشی، تکرار می‌گردد؛ بدین ترتیب مشاهدات مربوط به هر حالت جمع‌آوری می‌شوند. کل مشاهدات در هر حالت به M بخش (به تعداد تلفیق‌ها) تقسیم شده و از هر بخش میانگین‌گیری می‌شود. حاصل میانگین‌گیری به عنوان بردار میانگین توزیع اولیه‌ی تلفیق محسوب می‌گردد.

در شکل (۳) چگونگی مقداردهی اولیه اجزاء هر گروه PSO نشان داده می‌شود.



شکل(۳): مقداردهی اولیه اجزاء گروه

همانطور که در شکل مشاهده می‌شود، مجموعه نمونه‌های آموزشی به تعداد اجزاء مورد نظر برای تشکیل یک گروه تقسیم می‌شود و سپس مجموعه‌های کوچکتر، ورودی الگوریتم آموزشی مبتنی بر ML قرار می‌گیرد. خروجی الگوریتم آموزش فایلی است که در آن مشخصات مدل‌های مخفی مارکوف متناظر با تمامی واحدهای آوایی، بدست آمده است. لازم به ذکر است که آموزش مبتنی بر ML برای هر یک از مدل‌ها تا ۵ تکرار صورت می‌پذیرد و این در حالتی است که آموزش کامل مدل‌ها (تا همگرایی الگوریتم آموزش) و دادگان مشابه (نمونه‌های آموزشی موجود) در ۱۰ تکرار انجام می‌شود. به عبارت دیگر مدل‌های بدست آمده برای مقداردهی اولیه اجزاء گروه، مدل‌های خامی هستند که تنها برای سرعت بخشیدن به فرآیند جستجو مورد استفاده قرار می‌گیرند. مقداردهی اولیه بردارهای سرعت با اعداد تصادفی در بازه تعریف شده معین انجام می‌شود.

۴-۲- ارزیابی اجزاء

کیفیت هر جزء در فضای جستجو پس از هر حرکت، با استفاده از یک تابع کیفیت سنجیده می‌شود. به این منظور توابع متفاوتی در این تحقیق مورد استفاده قرار گرفته‌اند که در ادامه به تشریح آنها پرداخته می‌شود. با استفاده از تابع (۱۱) تعداد نمونه‌هایی که به صورت نادرست بازشناسی شده‌اند، ملاک عمل قرار گرفته و حین فرآیند آموزش HMM سعی می‌شود که تعداد نمونه‌هایی که اشتباه شناسایی می‌شوند، حداقل گردد.

$$\text{تابع کیفیت} = \frac{1}{\text{تعداد نمونه‌هایی که بصورت نادرست شناسایی می‌شوند}} \quad (۱۱)$$

تابع کیفیت خطای طبقه‌بندی به صورت رابطه (۱۲) در نظر گرفته شده است. با استفاده از این تابع، ارزیابی هر جزء (مدل HMM) با استفاده از میزان خطای طبقه‌بندی صورت می‌پذیرد.

$$\text{تابع کیفیت} = \frac{\text{میزان شباهت مدل به داده‌های آموزشی واحد بازشناسی متناظر}}{\text{میزان فاصله مدل از داده‌های آموزشی سایر واحدهای بازشناسی}} \quad (۱۲)$$

به بیان دیگر

$$\text{تابع کیفیت} = \frac{\sum_{i=1}^{w} g(O^i, \lambda_i)}{\sum_{j=1,,,w,\, j\neq i} g(O^i, \lambda j)} \quad (۱۳)$$

که در آن $\{O^i, i=1,2,\ldots,w\}$ داده‌های آموزشی و O^i خاص HMM، i ام برچسب گذاری شده است. $g(O^i, \lambda_i)$ شباهت شرطی O^i به مدل λ_i است. در حین آموزش، سعی بر آن است که مخرج کسر (خطای بازشناسی) حداقل و صورت کسر (قدرت جزء در بازشناسی واحد آوایی متناظر) حداکثر گردد.

انتظار می‌رود که مدل بهینه برای بازشناسی واحد آوایی، با حرکت اجزای گروه، به صورت متمایز از سایر مدلهای واحدهای آوایی دیگر، بدست آید. به عنوان مثال اگر مدلی درصد بازشناسی بالایی را برای واحد آوایی متناظر نتیجه دهد، اما در کنار سایر مدلها، بخوبی عمل نکند، به عنوان مدلی با کیفیت بالا در نظر گرفته نمی‌شود، زیرا حافظ تمایز میان مدل‌ها نیست. از دید طبیعی، به علت تعریف تابع کیفیت بصورت تمایزی(حالت تمایزی gbest و pbest) در آسمانی که انواع پرنده در حال پروازند، هر پرنده طوری حرکت می‌کند که احتمال اشتباه و حرکت کردن در پی یا در کنار انواع پرندگان دیگر کاهش یابد. نتایج بکارگیری این روش‌ها، همچنین استفاده از برخی معیارهای آموزش تمایزی، در بخش بعد مورد بررسی قرار گرفته است.

۵- نتایج آزمایشات و ارزیابی سیستم بازشناسی

در این بخش به نتایج تجربی استفاده از الگوریتم PSO در آموزش مدل مخفی مارکوف و تاثیر این روش در بهبود دقت بازشناسی گفتار براساس چارچوب تئوری آورده شده، پرداخته می‌شود. نتایج با استفاده از درصد دقت بازشناسی نمونه‌های موجود در مجموعه آزمون، همچنین میزان کاهش خطای بازشناسی روش بکار گرفته شده در مقایسه با الگوریتم k-means بیان می‌شود.

در ارزیابی عملکرد PSO، ۵۰۰ نمونه جمله آموزشی و ۱۰۰ نمونه جمله آزمون از دادگان فارس‌دات مورد استفاده قرار گرفته است.

در این تحقیق هر HMM مدل کننده یک واج است. در تمامی آزمایشات برای بازشناسی ۳۰ واج زبان فارسی، ۳۰ گروه در نظر گرفته شده است. شرط خاتمه اجرای تکرار الگوریتم‌ها، تعداد دفعات اجرا است. در پایان هر تکرار، مدل‌های مخفی مارکوف بدست آمده در فایلی ذخیره و نگهداری می‌شوند؛ سپس در ارزیابی مجموعه آزمون مورد استفاده قرار می‌گیرند. مدلهایی که بهترین نتیجه بازشناسی را حاصل کنند، به عنوان خروجی الگوریتم معرفی می‌گردند.

مدل‌های مخفی مارکوف سه حالته و هر حالت بصورت ۳ گوسی ۲۴ بعدی(شامل ۱۲ ضریب کپسترال و ۱۲ مشتق اول کپسترال)، در نظر گرفته شده است. همچنین هر گروه در PSO شامل ۱۲ جزء است.

۵-۱- بررسی حذف اثر pbest (حذف حافظه هر جزء)

در این سری از آزمایشات، مقدار c1 برابر با صفر قرار گرفته است. در این حالت هر جزء تنها با تاثیر از بهینه سراسری در فضای جستجو حرکت می‌کند. دقت بازشناسی مدل‌های بدست آمده در این آزمایش در جدول (۱) آمده است.

جدول(۱): اثر حذف حافظه هر جزء در دقت بازشناسی نهایی

ردیف	پارامترها	درصد بازشناسی مجموعه آزمون		درصد کاهش خطا بازشناسی
		PSO	K-means	
۱	Cv1=1وC1=0 Cv2=1وC2=2 Wvt=0.8وWt=1	۷۰،۱۳	۶۹،۴۱	۲،۴
۲	Cv1=1وC1=0.5 Cv2=1وC2=2 Wvt=0.8وWt=1	۶۹،۵۱	۶۹،۴۱	۰،۳
۳	Cv1=0وC1=0 Cv2=1وC2=1 Wvt=0.8وWt=1 حذف تابع rand()	۶۹،۴۱	۶۹،۴۱	۰

در هر سه آزمایش سرعت همگرایی نسبت به سایر آزمایشات بیشتر است، اما احتمال کمتری برای رسیدن به بهینه سراسری وجود دارد. در آزمایش شماره (۳) یک نسخه صددرصد اجتماعی قطعی از PSO با حذف ضرایب تصادفی اجرا می‌شود که بهبودی را در روند بازشناسی حاصل نمی‌کند.

۵-۲- بکارگیری معیارهای ارزیابی گوناگون

در آزمایشات این بخش تعداد نمونه‌های ناصحیح بازشناسی شده، ملاک عمل قرار گرفته و در مواردی که دو HMM تعداد برابر نمونه را اشتباه بازشناسی می‌نمایند، جهت بالا بردن دقت از امتیازی که هر مدل با استفاده از معیارهای تمایزی کسب می کند، استفاده می‌شود. نتایج در جدول (۲) آمده است.

جدول(۲): دقت بازشناسی واجها با استفاده از توابع کیفیت متفاوت

ردیف	معیار ارزیابی	درصد بازشناسی مجموعه آزمون		
		PSO	k-means	درصد کاهش خطا
۱	نتیجه بازشناسی+ امتیاز مدل در بازشناسی واجهای نامتناظر	۷۰،۰۳	۶۹،۴۱	۲،۰۶
۲	نتیجه بازشناسی+ معیار MCE	۷۰،۴۲	۶۹،۴۱	۳،۳۶
۳	نتیجه بازشناسی+معیار MMD	۶۹،۴۴	۶۹،۴۱	۰،۱
۴	نتیجه بازشناسی+ معیار MMI	۶۹،۴۴	۶۹،۴۱	۰،۱

آزمایش (۳) و (۴) با استفاده از معیار ارزیابی نتیجه بازشناسی و به همراه آن امتیاز معیار MMD یا MMI بهبود دقت قابل توجهی را نسبت به الگوریتم k-means حاصل نمی‌نماید. بکارگیری معیار MCE نتیجه بهتری را نسبت به سایر آزمایشات نشان می‌دهد.

۵-۳- ارزیابی زمانی PSO

نتایج ارزیابی زمانی فرآیند آموزش در جدول (۳) نمایش داده می‌شود.

جدول (۳): ارزیابی زمان الگوریتم PSO

متوسط زمان لازم	بر حسب دقیقه
یک حرکت ۱۶ جزئی برای تنها یک مدل HMM	۴
ارزیابی یک جزء گروه	۱۲ ثانیه
یک حرکت ۱۶ جزئی برای کلیه مدلهای HMM	۱۲۸

یک حرکت کامل گروهی با ۱۶ جزء و سپس ارزیابی کیفیت اجزاء آن، حدود ۴ دقیقه بطول می‌انجامد. زمان لازم برای ارزیابی یک جزء نیز ۱۲ ثانیه است؛ پس بطور متوسط ۱۲*۱۶ ثانیه، برابر با ۳/۲ دقیقه مختص به ارزیابی اجزاء و ۰/۸دقیقه جهت حرکت اجزاء سپری می‌شود. بطور متوسط تعداد تکرارهای مورد نیاز برای الگوریتم PSO جهت همگرا شدن ۱۵ تا ۲۰ تکرار است و هر تکرار PSO، ۱۲۸ دقیقه بطول می‌انجامد، در شرایطی که آموزش کامل مدلهای HMM با استفاده از الگوریتم k-means به ۴ دقیقه زمان نیاز دارد. اشکال استفاده از روش PSO جهت آموزش مدل مخفی مارکوف نسبت به روشهای مبتنی بر ML، پیچیدگی زمانی بالاتر آن است. اما به دلیل اینکه فرآیند آموزش HMM، قبل از بازشناسی انجام می‌شود، بازشناسی بی‌درنگ را تحت تاثیر قرار نمی‌دهد.

۶- جمع بندی و راه‌کارهای آتی

الگوریتم PSO همچنین نحوه بکارگیری آن در فرآیند آموزش مدل مخفی مارکوف تشریح شد. با در نظر داشتن قابلیت الگوریتم PSO در جستجوی بهینه سراسری و نیز با کمک گرفتن از معیارهای آموزش تمایزی MCE، MMD و MMI در روند آموزش، پارامترهای مدل مخفی مارکوف پیوسته یک سیستم بازشناسی گفتار فارسی به دست آورده شد. استفاده از معیار تمایزی MCE به همراه نتیجه بازشناسی، جهت سنجش کیفیت آموزش مدل‌ها، بهترین دقت بازشناسی را نسبت به بکارگیری سایر ملاک‌های سنجش حاصل می‌نماید و از کارایی آموزش بالاتری برخوردار است. در این شرایط، درصد خطای مدل‌های مخفی مارکوف به دست آمده با استفاده از الگوریتم PSO در بازشناسی هر واج متعلق به مجموعه آزمون، نسبت به درصد خطای بازشناسی مدل‌های بدست آمده از آموزش الگوریتم k-means قطعه‌ای، بطور متوسط ۳،۳۶ درصد کاهش یافته است.

با توجه به آن که حرکت، تنها عملگر تعریف شده در PSO است، آنچه که می‌تواند در ادامه کار انجام پذیرد، استفاده از دانش مساله و بکارگیری مکاشفه‌هایی مانند تعریف عملگرهای جدید، خاص مدل‌های مخفی مارکوف است، تا به این وسیله دقت بازشناسی افزایش و یا سرعت همگرایی الگوریتم PSO و در نتیجه زمان لازم برای آموزش مدل‌های مخفی مارکوف کاهش یابد.

سپاسگزاری

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام شده است.

مراجع

[۱] هدیه ساجدی، حسین ثامتی، حمید بیگی، باقر باباعلی"**کاهش خطای طبقه بندی مدل مخفی مارکوف در بازشناسی گفتار فارسی با بکارگیری الگوریتم ژنتیکی در روند آموزش**"، کنفرانس فنآوری اطلاعات،۱۳۸۴

[۲] هدیه ساجدی، حسین ثامتی، باقر باباعلی"**بهبود دقت بازشناسی گفتار فارسی با بکارگیری الگوریتم ژنتیکی در روند آموزش تمایزی مدل مخفی مارکوف**" ، کارگاه آموزشی خط و زبان فارسی و رایانه، ص۴۴۸-۴۵۶، ۱۳۸۵

[3] R.Bahl,P.F.Brown.P.V.de Souza,R.L.Mercer,"*Maximum Mutual Information estimation of hidden Markov model parameters for speech recognition*", Proc IEEE, pp. 49-52,1986

[4] Eberhart, R.C. and Shi, Y., "*Evolving artificial neural networks*", Proceedings of International Conference on Neural Networks and Brain, 1998, Beijing, P.R. China, pp. PL5-PL13, 1998

[5] R.Bahl,P.F.Brown.P.V.de Souza,R.L.Mercer,"*Maximum mutual information estimation of hidden Morkov model parameters for speech recognition*",Proc IEEE,pp. 49-52,1986

[6] Ephraim Y.,Dembo A.,Rabiner L.R., "*A Minimum Discrimination Information Approach for Hidden Markov Modeling*",IEEE,Trans,vol.35,no.5,pp. 1001-1013,1989.

[7] Eprahim Y.,Dembo A.,Rabiner L.R., "*A Minimum Discrimination Information Approach for Hidden Markov Modeling*",IEEE,ICASSP'87,1987

[8] S.Kwong,Q.H. He, K.F. Man,K.S.Tang, "*A maximum model distance approach for HMM-based speech recognition*",Pattern Recognition ,vol 31,No.3,pp,219-229,1998

[9] B.H.Juang,Wu Chou,Chin-Hui Lee, "*Minimum classification error rate methods for speech recognition*

".IEEE Trans.Speech Audio Process,vol 5,No.3,pp,257-265,1997

[10] K.E.Parsopoulos,M.N.Vrahatis, "*Recent approaches to global optimization problems through Particle Swarm Optimization*",Natural Computing,pp.235-306,2002

[11] Kennedy, J. and Eberhart, R.C., "*Particle swarm optimization*", Proc. IEEE int'l conf. on neural networks, Vol. IV, pp. 1942-1948, IEEE service center, Piscataway, NJ, 1995

[12] J.Kennedy,R.C.Eberhart, "*Particle Swarm Optimization*", IEEE Conference on Neural Networks,Vol.4,PP.1942-1948,1995

[13] Clerc, Kennedy, "*The particle swarm Explosion,stability and convergence in a multideimentional complex space*", IEEE Transaction on Evolutionary Computation,pp.58-73,2002

[14] Eberhart RC,Simpson P,Dobbins R, "*Computational Intelligence PC Tools*",Academic Press,1996

[15] Gudise, V.G., and Venayagamoorthy, G.K., "*Comparison of particle swarm optimization and back propagation as training algorithms for neural networks*", Proceedings of the IEEE Swarm Intelligence Symposium 2003 (SIS 2003), Indianapolis, Indiana, USA, pp. 110-117, 2003

[16] R.Eberhart,J.Kennedy, "*New Optimizer using particle swarm theory*",Micro Machine and Human Science,pp.39-43,1995

[17] Morten Lovbjerg, "*Improving Particle Swarm Optimization by hybridization of stochastic search heuristics and Self-Organizad Criticality*",Supervisor: Thiemo Krink, University of Aarhus,May 2002

---

[1] Maximum likelihood

[2] Topology

[3] Maximum Mutual Information

[4] Corrective Training

[5] Particle swarm optimization

[6] Swarm

[7] Particle

[8] State

[9] Minimum Discrimination Information

[10] Maximum Model Distance

[11] Minimum Classification Error

[12] Swarm Systems

[13] Artificial Life

[14] Back Propagation