تشخیص حالت دست با استفاده از میدان تصادفی شرطی پنهان پنجره‌ای با پارامترهای غیرمشترک

مصطفی رفیعی[1]، سید کمال‌الدین غیاثی شیرازی[2]، احد هراتی[3]

[1] دانشگاه فردوسی مشهد ، mostafa.rafie@stu.um.ac.ir

[2] دانشگاه فردوسی مشهد ، k.ghiasi@um.ac.ir

[3] دانشگاه فردوسی مشهد ، a.harati@um.ac.ir

چکیده

مدل‌های گرافی احتمالاتی با لایه مخفی چارچوبی قدرتمند برای دسته‌بندی دنباله‌ای از داده‌ها ارائه می‌کنند. میدان تصادفی شرطی با حالات پنهان(HCRF) از جمله مدل‌های تفکیکی است که از لایه مخفی استفاده کرده و وزن‌های مشترکی را برای تمامی قاب‌ها یاد می‌گیرد. در این مقاله با الهام از HCRF مدلی جدید معرفی می‌کنیم که در پنجره‌ای از متغیرهای مخفی وزن‌های مجزا (غیرمشترک) برای هر متغیر مخفی یاد می‌گیرد. همچنین در مدل پیشنهادی، بر خلاف HCRF، تعداد حالات مخفی برای هرکدام از متغیرهای پنهان عضو پنجره می‌تواند متفاوت انتخاب شود. این کار با کاهش فضای حالت جستجو برای متغیرهای مخفی سرعت و کیفیت استنتاج را بهبود داده و امکان انتساب ناهمسان اهمیت به تک فریم‌ها را فراهم می‌آورد. در عین حال ما با اتصال زنجیروار خروجی پنجره‌های متوالی امکان نشر اطلاعات بین برچسب‌های هر پنجره را فراهم می‌آوریم. آزمایش‌های ما بر روی کاربرد تشخیص حالت دست صورت گرفته که نشان‌دهنده عملکرد بهتر مدل پیشنهادی نسبت به سایر مدل‌های گرافی احتمالاتی در این حوزه است.

واژه‌های کلیدی

پارامترهای غیرمشترک، میدان تصادفی شرطی با حالات پنهان پنجره‌ای (WHCRF)، حالات مخفی، تشخیص حالت حرکت

۱- مقدمه

یکی از مسائل مهم در دنیای واقعی برچسب‌دهی به دنباله‌ای از داده‌ها است. در این گونه مسائل با استفاده از مشاهده دنباله‌ای از قاب‌های پشت سرهم به دنبال یافتن برچسب متناظر با هرکدام از قاب‌ها هستیم. از جمله کاربردهای مهم در این زمینه می‌توان تشخیص حالات حرکت را نام برد که در واقع به دنبال تفسیر نمادهایی است که با استفاده از حرکات بدن یا بخشی از آن مثل دست ارائه می‌شوند. انجام تحقیقات بر روی کاربرد تشخیص حالات حرکت می‌تواند برای درک زبان بدن انسان توسط رایانه‌ها مفید واقع شود. برای حل مسائلی از این دست، عموماً از مدل‌های گرافی احتمالاتی استفاده می‌شود. این مدل‌ها خود به دو گروه مدل‌های مولد[1] و تفکیکی[2] تقسیم‌بندی می‌شوند. در مدل‌های مولد به دنبال بازنمایی توزیع مشترک بین ورودی و خروجی هستیم. از این رو در این مدل‌ها نیاز به مدل کردن توزیع احتمال پیشین ورودی وجود دارد. از جمله مدل‌های مهم در این حوزه مدل پنهان مارکوف[3] و میدان تصادفی مارکوف[4] را می‌توان نام برد. در برخی از کاربردها مدل کردن داده های ورودی و بازنمایی توزیع بر روی آنها کاری سخت و پیچیده است. از این رو مدل‌های تفکیکی در این زمینه مفید به نظر می‌رسند زیرا در این مدل‌ها با توجه به شرطی شدن بر روی داده های ورودی دیگر نیازی به مدل‌سازی و بازنمایی آنها وجود ندارد. در مدل‌های تفکیکی به دنبال بازنمایی توزیع شرطی بر روی داده‌های ورودی هستیم.

مدل‌های گرافی احتمالاتی را همچنین می‌توان از حیث داشتن یا نداشتن لایه مخفی مورد دسته‌بندی‌های ریزتری قرار داد. برای مثال مدل پنهان مارکوف دارای یک لایه مخفی است که در تشخیص ساختارهای پیچیده به آن کمک می‌کند. این مدل یک مدل مارکوف آماری است که در آن سامانه مدل شده به صورت یک فرایند مارکوف با حالت‌های پنهان فرض می‌شود. میدان تصادفی مارکوف، مدل مولد دیگری است که لایه مخفی ندارد. این مدل نمونه‌ای از یک مدل آماری است که ارتباط و استقلال بین متغیرهای تصادفی را به کمک یک گراف ساده و بدون جهت مدل می‌کند[10].

---

[1] Generative

[2] Discriminative

[3] Hidden Markov Model (HMM)

[4] Markov Random Fields (MRF)

در مقابل مدل‌های گرافی احتمالاتی تفکیکی قدرتمند میدان تصادفی شرطی[5] و میدان تصادفی شرطی با حالات پنهان که توزیع احتمال برچسب به شرط مشاهدات را به صورت مستقیم مدل‌سازی کرده و بنابراین، نواقص مدل‌های مولد را، که توزیع مشترک برچسب و مشاهدات را با فرض مستقل بودن مشاهدات به شرط برچسب محاسبه می‌کنند، دارا نمی‌باشند. در میدان تصادفی شرطی با توجه به نداشتن لایه مخفی نمیتوانند به خوبی ساختارهای درونی بین داده‌های ورودی را یاد بگیرند. در ساختار این مدل بین ورودی و برچسب‌ها رابطه میانی دیگری وجود ندارد که این امر موجب عدم استخراج کافی اطلاعات از ساختار داده‌ها می‌شود[2] .

مدل میدان تصادفی شرطی با حالت پنهان[6] با اضافه کردن لایه مخفی به مدل CRF به دنبال استخراج روابط بین ساختارهای ورودی و در نتیجه عملکرد بهتر نسبت به آن بود[4]. این مدل توزیع مشترک بین حالات مخفی و برچسب‌ها را با شرط مشاهده نمودن ورودی‌ها بازنمایی می‌کند.

در این مقاله ما مدل HCRF پنجره‌ای با پارامترهای غیرمشترک را ارائه می‌دهیم که با استفاده از پنجره‌ای که بر روی حالات مخفی تعریف می‌کند می‌تواند برای هرکدام از قاب‌های پنجره وزن‌های متفاوتی را یاد بگیرد. مدل HCRF یک الگو برای دنباله‌ای از داده‌ها معرفی می‌کند که این الگو برای تمام قاب‌های دنباله تکرار می‌شود. منظور از این الگو همان وزن‌های مدل هستند که در مرحله آموزش یاد گرفته شده‌اند. این وزن‌ها نشان‌دهنده میزان وابستگی ارتباطات متغیرهای مختلف به هم هستند. مدل HCRF این وزن‌ها را برای قاب‌های کل دنباله تکرار می‌کند یا به عبارت دیگر وزن‌ها در مدل به اشتراک گذاشته می‌شوند. مدل پیشنهادی ما بر روی پنجره‌ای با اندازه ثابت کار می‌کند که بر روی دنباله ورودی لغزانده می‌شود تا تمامی بخش‌های آن‌را پوشش دهد. اندازه پنجره با توجه به کاربرد مورد نظر توسط طراح مدل مشخص می‌شود. بدین صورت برای هر قاب از ورودی واقع شده در پنجره، وزن‌های متفاوتی یاد گرفته می‌شود. با توجه به حرکت پنجره لغزان با اندازه ثابت بر روی داده‌ها، مدل همواره با دنباله‌های با طول مشخص (همان اندازه پنجره) آموزش می‌بیند و لذا یادگیری پارامترهای غیرمشترک ممکن می‌شود. از طرفی در مرحله آزمون این قابلیت وجود دارد که در کاربردهایی که تنها به یک برچسب برای کل دنباله ورودی نیاز است، مدل بر روی کل دنباله استنتاج انجام دهد و به دقت بالاتری دست یابد. در عین حال همیشه پردازش ابتدایی در سطح پنجره انجام می‌شود. به بیان دیگر در مرحله آزمون، پنجره تعریف شده بر روی حالات مخفی بر روی دنباله ورودی با گام یک قاب حرکت داده شده و برچسب‌های هر کدام از این پنجره‌ها در لایه جدیدی به هم مرتبط می‌شوند. در واقع در این لایه جدید سازگاری برچسب‌ها با هم مورد پردازش قرار می‌گیرد.

همچنین ما برای هر قاب در پنجره تعریف شده می‌توانیم تعداد حالات مخفی را به صورت دستی انتخاب نماییم که با این کار فضای حالات مخفی کاهش پیدا می‌کند اما در مدل HCRF برای تمام فریم‌های دنباله تعداد حالات مخفی ثابت در نظر گرفته می‌شود. ما در مدل پیشنهادی برای فریم‌های میانی با توجه به اینکه اهمیت بیشتری دارند تعداد حالات بیشتری در نظر می‌گیریم.

۲- کارهای مرتبط

از جمله اولین مدل‌های گرافی احتمالاتی که در حوزه تشخیص حرکت انسان مورد استفاده قرار گرفت،مدل پنهان مارکوف بود. از این مدل برای تشخیص حرکت سر استفاده شد[1] . در همان سال مدل تفکیکی CRF معرفی شد که برای دنباله‌ای از داده‌ها مورد استفاده قرار گرفت[2] . چند سال بعد از این مدل برای تشخیص حرکت متنی[7] انسان استفاده گردید و توانست نسبت به مدل پنهان مارکوف عملکرد بهتری داشته باشد[3] .

در ادامه تعمیمی از مدل CRF با نام CRF پویا معرفی شد که ساختار و پارامترهای را بر روی مدل تکرار می‌کند و با این کار می‌تواند روابط پیچیده‌تر بین داده‌های ورودی را بهتر مدل کند[17,16] .

در سال 2007 با اضافه کردن لایه مخفی به مدل CRF مدل جدید HCRF ارائه شد که با استفاده از حالات مخفی زیرساختارهای درونی و بیرونی داده‌های ورودی را تشخیص داده و در نتیجه استخراج اطلاعات برای شناسایی برچسب دنباله ورودی را بهتر انجام می‌دهد. از این مدل برای داده‌های تقطیع شده استفاده می‌شود که در آن به هر دنباله تنها یک برچسب تعلق می‌گیرد. این مدل برای اولین بار در حوزه تشخیص حالات حرکت استفاده گردید[4] . در همان زمان مدل دیگری ارائه شد به نام میدان تصادفی شرطی پویای پنهان[8] که برای دنباله‌های دارای چند برچسب مورد استفاده قرار می‌گرفت[5] .

بر خلاف نگاه متداول ستنی در مدل‌های احتمالاتی که تمامی داده مشاده شده در هر حالت را در یک متغیر واحد تحت نام *مشاهده* جمع می‌کند، می‌توان هر ویژگی قابل استخراج از داده‌های خام را به صورت یک متغیر مشاهده‌شده جدید و جدا از سایر مشاهدات به مدل اضافه کرد[18] . در کار یل سونگ و همکاران [6] ، این نگرش با درنظر گرفتن ویژگی‌های استخراج شده از ورودی خام دنبال شده و برای هر قاب چند متغیر ورودی در مدل منظور شده است. آن‌ها توپولوژی‌های مختلفی را برای ارتباط بین حالات مخفی و ورودی مطالعه کرده و مدل‌های HCRF متصل[9] و HCRF تزویجی[10] را معرفی می‌کنند. همچنین در این مقاله گونه‌های جدیدی نیز برای میدان تصادفی شرطی پویای پنهان معرفی می‌شود.

برای استخراج روابط غیرخطی و پیچیده دنباله داده‌های ورودی می‌توان از شبکه عصبی در مدل‌های تفکیکی کمک گرفت. در سال 2009 مدل میدان عصبی شرطی[11] ارائه شد که بین لایه ورودی و خروجی مدل CRF یک لایه شامل تعدادی توابع گیت[12] اضافه می‌کند. همچنین در این مقاله پارامترهای شبکه عصبی همراه با پارامترهای مدل آموزش داده می‌شود[15]

---

[5] Conditional Random Fields (CRF)

[6] Hidden-state Conditional Random Fields (HCRF)

[7] contextual

[8] Latent Dynamic Conditional Random Fields (LDCRF)

[9] Linked-HCRF(LHCRF)

[10] Coupled-HCRF(CHCRF)

[11] Conditional Neural Fields (CNF)

[12] Gate Function

. پس از این برای مدل HCRF نیز از این روش استفاده شد که در آن لایه مورد نظر بین ورودی و لایه مخفی قرار می‌گیرد[19] .

یادگیری سلسله مراتبی از جمله روش‌هایی است که در حوزه مدل‌های گرافی نیز مورد بررسی قرار گرفته است. در مدل تفکیکی CRF یادگیری سلسله مراتبی استفاده می‌شود. در این مدل‌ها هدف اصلی استفاده از لایه پایین مدل برای ساخت لایه بالاتر است[21,20] . مدلCRF با ساختار عمیق[13] همانند دیگر روش‌های سلسله مراتبی دارای چندین لایه ساده CRF است تا بتواند اطلاعات بیشتری از روابط بین دادگان ورودی استخراج نماید. اما هدف اصلی این مدل، یادگیری لایه‌های میانی با استفاده از داده-های ورودی و ترکیب همه این اطلاعات با هم است. این کار را با پیش‌آموزش بدون ناظر لایه به لایه مبتنی بر آنتروپی و تنظیم دقیق مبتنی بر پس‌انتشار درست‌نمایی انجام می‌دهد[14] . یک سال بعد همین روش برای مدل HCRF استفاده گردید و نام آن HCRF با ساختار عمیق گذاشته شد که لایه متغیرهای مخفی به لایه پایانی مدل قبل اضافه شده است[13].

یل سونگ و همکاران[11] نیز یک روش سلسله مراتبی برای نمایش ویژگی-های ورودی در نظر گرفتند که در هر مرحله ویژگی‌های ورودی را به گونه‌ای خلاصه‌سازی می‌کند که مشاهدات شبیه هم را به جای اینکه در چند قاب نشان دهد، در یک قاب آنها را نمایش می‌دهد. برای یادگیری دنباله‌ها در هر مرحله از HCRF استفاده شده که از متغیرهای مخفی آن برای نمایش دنباله جدید استفاده می‌شود و در هر مرحله خلاصه‌ای از ویژگی‌های ورودی مرحله قبل یاد گرفته می‌شود. در ادامه نیز یک روش فیلتر چندلایه با یک پنجره زمانی برای برچسب‌دهی و تقطیع دنباله‌ها ارائه گردید و از آن برای تشخیص حرکت استفاده شد[24].

اخیرا از شبکه‌های عصبی پیچشی[14] برای ترکیب با مدل‌های گرافی استفاده گردیده است. در این تحقیق یک مدل مکانی-زمانی[15] شبکه عصبی پیچشی جدید برای ترکیب با میدان تصادفی شرطی استفاده شده است. این مدل از یک شبکه عصبی پیچشی جدید برای استخراج ویژگی‌های مناسب از ورودی خام استفاده می‌کند و سپس از یک میدان تصادفی شرطی برای پیش‌بینی برچسب‌های هر قاب استفاده می‌کند. از این‌رو نام این مدل میدان تصادفی عصبی پیچشی[16] نامیده شد[23].

در تعدادی دیگر از کارها با توجه به اینکه استفاده از داده‌ها با ابعاد بالا نیاز به داده‌های آموزشی زیاد برای جلوگیری از بیش‌برازش[17] دارد می‌توان با کاهش ابعاد ویژگی‌های خام ورودی از داده‌های آموزشی کمتری استفاده نمود. برای کاهش کافی ابعاد داده‌های ورودی با استفاده از روش کاهش ابعاد دنباله با کرنل[18] این کار را انجام داد. در این روش دیگر ما فرضیات قوی بر روی توزیع داده‌های ورودی نداریم[12].

۳-پیش‌زمینه

فرض کنیم می‌خواهیم برچسب‌های y را از ورودی $\mathbf{x}$ شناسایی نماییم. هر برچسب y عضو مجموعه Y است که شامل همه برچسب‌های ممکن است و هر بردار x یک بردار مشاهدات $\boldsymbol{x} = \{x_1 \dots x_m\}$ است. هرکدام از x_j به‌وسیله یک بردار ویژگی d بعدی نشان داده می‌شود. ما برای هر نمونه ورودی $\mathbf{x}$ یک بردار متغیرهای پنهان $\boldsymbol{h} = \{h_1 \dots h_m\}$ فرض می‌کنیم و هر h_j متناظر با یک برچسب‌دهی x_j است. با استفاده از تعریف‌های انجام‌شده می‌توان مدل احتمالاتی شرطی با متغیرهای پنهان را به‌صورت زیر تعریف نمود:

$$P(y, \boldsymbol{h} \mid \boldsymbol{x}, \theta) = \frac{e^{\psi(y, \boldsymbol{h}, \boldsymbol{x}; \theta)}}{\sum_{y', \boldsymbol{h}} e^{\psi(y', \boldsymbol{h}, \boldsymbol{x}; \theta)}} \quad (1)$$

که $\theta = \{\alpha, \beta\}$ پارامترهای مدل و $\psi(y, \boldsymbol{h}, \boldsymbol{x}; \theta) \in \Re$ تابع پتانسیل است. پارامتر α نشان‌دهنده پتانسیل‌های مرتبه یک و β نشان‌دهنده پتانسیل‌های مرتبه دو است.

در مدل HCRF برای کل دنباله یک برچسب انتخاب می‌شود و وزن‌های آموزش داده شده برای هر قاب تکرار می‌شوند. به عبارتی وزن‌های بین حالات مخفی و ورودی و خروجی برای تمام قاب‌ها به اشتراک گذاشته شده‌اند.در شکل 1 مدل گرافی HCRF را مشاهده می‌کنید. در مدل HCRF تعداد حالات مخفی در هر قاب با هم برابر بوده و همچنین وزن‌ها در هر قاب تکرار می‌شوند. همانطور که در شکل 1 مشاهده می‌کنید در این مدل سه نوع پتانسیل وجود دارد. می‌توان توابع پتانسیل به صورت دو تابع ویژگی f و g معرفی کرد که به صورت زیر نمایش داده می‌شود:

$$\begin{aligned} &\psi(y, \boldsymbol{h}, \boldsymbol{x}; \theta) \\ &= \sum_{j=1}^{m} \sum_{l \in L_1} \alpha_l \, f_l(j, y, h_j, \boldsymbol{x}) \\ &+ \sum_{(j.k) \in E} \sum_{l \in L_2} \beta_l \, g_l(j, k, y, h_j, h_k, \boldsymbol{x}) \end{aligned} \quad (2)$$

که L_1 و L_2 مجموعه‌های به ترتیب پتانسیل‌های مرتبه یک و مرتبه دو بوده و E گراف تشکیل شده از متغیرها می‌باشد. بنابراین برای یافتن برچسب y برای کل دنباله با سرجمع‌گیری روی حالات مخفی داریم:

$$P(y \mid \boldsymbol{x}, \theta) = \sum_{\boldsymbol{h}} P(y, \boldsymbol{h} \mid \boldsymbol{x}, \theta) = \frac{\sum_{\boldsymbol{h}} e^{\psi(y, \boldsymbol{h}, \boldsymbol{x}; \theta)}}{\sum_{y', \boldsymbol{h}} e^{\psi(y', \boldsymbol{h}, \boldsymbol{x}; \theta)}} \quad (3)$$

---

[13] Deep-Structured CRF

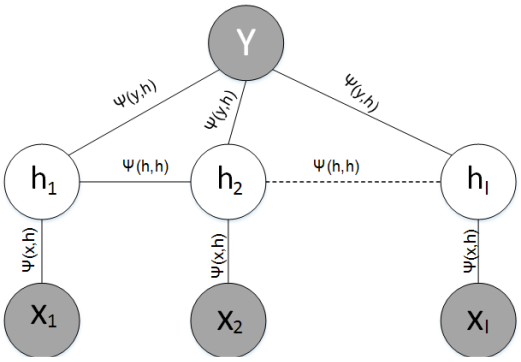
[14] convolutional neural network (CNN)

[15] spatio-temporal

[16] Convolutional Neural Random Field

[17] Overfitting

[18] Sequence Kernel Dimension Reduction



شکل 1 - درمدل HCRF وزن‌های مربوط به هر فاکتورمشخص شده در شکل در هر قاب تکرارمی‌شوند و تعداد حالات هر متغیر مخفی در دنباله یکسان است.

۴-مدل پیشنهادی

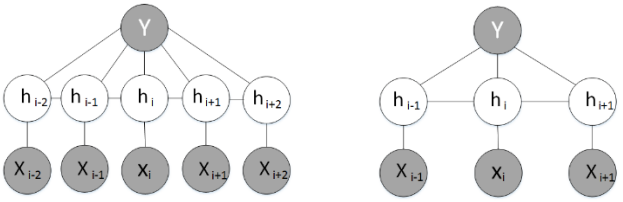
برای بیان بهتر مدل ما روش پیشنهادی را در سه بخش مدل، آموزش و استنتاج مورد بررسی قرار می‌دهیم.

۴-۱- مدل

در مدل پیشنهادی ما برای هر قاب در پنجره تعریف شده وزن‌های متفاوتی یاد می‌گیریم. به عبارتی وزنهای قاب‌های مختلف در پنجره مورد نظر دارای اشتراک نیستند. تعداد حالات مخفی برای هرکدام از قاب‌ها می‌توانند به صورت دستی انتخاب شوند به گونه‌ای که تعداد حالات در هر قاب می‌تواند تفاوت داشته باشد.

در شکل2 یک پنجره به طول 2s+1 در نظر گرفته شده است. همانطور که در شکل مشخص است هرکدام از وزن‌های فاکتورهای موجود با هم متفاوت هستند. همچینین در پنجره مزبور هرکدام از متغیرهای مخفی تعداد حالات متفاوتی دارند. شکل3 دو مثال از پنجره مدل را نشان داده که اندازه آن را s=1 و s=3 در نظر گرفته شده است. ما این پنجره را بر روی دنباله‌ای از مشاهدات حرکت می‌دهیم. در شکل4 مدل گرافی استفاده شده را مشاهده می‌کنید. در این شکل پنجره بر روی دنباله‌ای از مشاهدات حرکت کرده و پنجره‌ها می‌توانند از طریق برچسب‌های خود با پنجره‌های مجاور در ارتباط باشند. در کاربرد مورد مطالعه در مقاله حاضر، در نهایت یک برچسب با استفاده از برچسب‌های هرکدام از پنجره‌ها به کل دنباله مشاهدات داده می‌شود.

نکته قابل توجه این است که در مرحله آموزش برچسب‌های هر پنجره را مشاهده شده در نظر گرفته که در نتیجه هر پنجره از سایر پنجره‌ها مستقل می‌شود. بنابراین ما در این مرحله هرکدام از پنجره‌ها به صورت جداگانه آموزش داده و فاکتور بین برچسب‌ها را به صورت دستی مقداردهی می‌کنیم. در مرحله آزمون برچسب‌ها مشاهده شده نیستند و ارتباطی که بین برچسب‌ها وجود دارد باعث نشر اطلاعات بین پنجره‌های مختلف شده و در تشخیص برچسب واقعی به مدل کمک می‌کند. برای مشخص کردن برچسب کل



شکل 3- دو نمونه از مدل گرافی HCRF پنجره‌ای با طول پنجره متفاوت: پنجره به طول 3 قاب در سمت راست و در سمت چپ پنجره به طول 5 قاب نمایش داده شده‌اند. گره‌های خاکستری متغیرهای مشاهده‌شده در هر مورد را نشان می‌دهند.

دنباله، احتمال بیشترین برچسب را با استفاده از برچسب هر پنجره بدست آورده و از آن برای برچسب کل دنباله استفاده می‌کنیم.

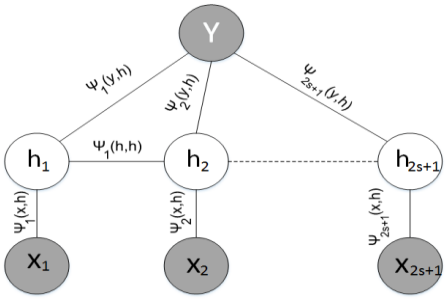
۴-۲- آموزش

برای آموزش مدل باید طول دنباله داده‌های آموزشی را به اندازه طول قاب‌های پنجره مدل تبدیل کنیم که برای این کار پنجره‌ای بر روی داده‌های ورودی حرکت داده که در هر مرحله یک به اندازه یک قاب جابه‌جا شده و در هر مرحله یک نمونه آموزشی به اندازه تعداد فریم‌های مدل تعریف می‌کند. ما مجموعه داده‌های آموزش $\mathcal{D} = \{(\boldsymbol{x_1}.\boldsymbol{y_1}).(\boldsymbol{x_2}.\boldsymbol{y_2}).\ldots.(\boldsymbol{x_n}.\boldsymbol{y_n})\}$ را در نظرمی‌گیریم که شامل n داده آموزشی برچسب‌دار است و هر $\boldsymbol{x_i}$ دارای n_i قاب است. حال برای یافتن پارامترهای بهینه همانند HCRF از بیشترین درست‌نمایی منظم شده استفاده می‌نماییم.

$$L(\theta) = -\sum_{i=1}^{n}\sum_{k=1}^{n_i-2s} log\, p(y_{i.k}|\boldsymbol{x}_{i.k},\theta) + \frac{\|\theta\|^2}{2\sigma^2} \qquad (4)$$

تابع هدف از دو عبارت تشکیل شده است که عبارت اول همان لگاریتم درستنمایی داده‌ها است و عبارت دوم جمله منظم سازی است که پاسخ‌های غیر منتظره را جریمه می‌کند. همچنین منظور از $\boldsymbol{x}_{i.k}$ نمونه‌های آموزش مدل با طول اندازه پنجره هستند.

برای محاسبه بردار گرادیان یک داده آموزشی خاص $\boldsymbol{x}_{i.k}$ نسبت به هرکدام از پارامترهای مدل باید مشتق گرفته شود.



شکل 2-مدل HCRF پنجره‌ای با جزئیات. وزن‌های مربوط به متغیرها در هر قاب غیرمشترک هستند و تعداد حالات در هر قاب پنجره می‌تواند متفاوت باشد و طول دنباله ثابت در نظر گرفته می‌شود.

$$L_{i.k}\left(\theta\right) = log\, p(y_{i.k}|\boldsymbol{x}_{i.k},\theta)$$
$$= log\left(\frac{\sum_{\boldsymbol{h}} e^{\psi(y,\boldsymbol{h},\boldsymbol{x};\theta)}}{\sum_{y',\boldsymbol{h}} e^{\psi(y',\boldsymbol{h},\boldsymbol{x};\theta)}}\right) \quad (5)$$

در اینجا با توجه به اینکه وزن‌های مدل اشتراکی ندارند پس تعداد پارامترهای مدل ما نسبت به HCRF به اندازه 2s+1 برابر بیشتر خواهد بود. پس به همان اندازه تعداد مشتق‌های مدل نیز افزایش می‌یابد. اگر نسبت به پارامترهای مربوط به پتانسیل‌های مرتبه یک مشتق بگیریم داریم:

$$\frac{\partial L_{i.k}\left(\theta\right)}{\partial \alpha_l} = \frac{\partial}{\partial \alpha_l}\log\left(\frac{\sum_h e^{\psi(y,h,x;\alpha)}}{\sum_{y',h} e^{\psi(y',h,x;\alpha)}}\right)$$
$$= \sum_{j.a} P\left(h_j = a|y,\boldsymbol{x},\alpha\right) \times f_l\left(j,y,h_j,\boldsymbol{x}\right) \quad (6)$$
$$- \sum_{y'}\sum_{\boldsymbol{j.a}} \left(P\left(h_j = a,y'|\boldsymbol{x},\alpha\right) \times f_l\left(j,y',h_j,\boldsymbol{x}\right)\right)$$

همچنین برای پتانسیل های مرتبه دو می‌توان نوشت:

$$\frac{\partial L_{i.k}(\theta)}{\partial \beta_l} = \sum_{(j.k)\in E.a.b} \left(P\left(h_j = a, h_k = b \mid y_i,\boldsymbol{x},\beta\right) g_l\left(j,k,y_i,a,b,\boldsymbol{x}\right)\right) \quad (7)$$
$$- \sum_{y'}\sum_{\boldsymbol{j,a}} \left(P(h_i = a, h_k = b, y'|\boldsymbol{x},\beta) g_l\left(j,y',a,b,h_j,\boldsymbol{x}\right)\right)$$

برای آموزش مدل ما برای هر داده آموزشی بردار گرادیان محاسبه کرده و بردارهای گردایان بدست آمده از هر داده آموزشی را با هم جمع می‌کنیم، به عبارت دیگر آموزش به صورت دسته‌ای است.

۳-۴- استنتاج

از استنتاج هم در مرحله آزمون و هم آموزش استفاده می‌شود. در مرحله آموزش برای محاسبه گرادیان نیاز به باورها یا همان احتمالات هر گره داریم، پس برای محاسبه گرادیان به استنتاج روی مدل نیاز است. در مرحله آزمون نیز برای بدست آوردن بهترین برچسب نیاز به استنتاج بر روی مدل داریم. با توجه به اینکه استنتاج در این دو مرحله با هم متفاوت است، در دو بخش به بیان آن می‌پردازیم.

۱-۳-۴-مرحله آموزش

از آنجا که در مرحله آموزش ارتباطی بین برچسب‌ها وجود ندارد، مدلی که بر روی آن استنتاج انجام می‌شود، هرکدام از پنجره‌های مدل است.

با توجه به شرطی‌سازی بر روی برچسب‌ها، ساختار گرافی حاصل به شکل زنجیر است که می‌توان از الگوریتم‌های استنتاج دقیق استفاده کرد. برای استنتاج بر روی این شبکه ما از الگوریتم پیش‌رو-پس‌رو استفاده می‌کنیم.
با توجه به اینکه برای هر قاب موجود در پنجره تعداد حالات می‌تواند متفاوت باشد، پیاده‌سازی پیش‌رو-پس‌رو برای مدل HCRF را به گونه‌ای تغییر دادیم تا بتواند عملیات استنتاج را بر روی مدل جدید با تعداد حالات متفاوت برای هر قاب انجام دهد. از این رو تفاوت در سایز پیام های ارسالی بین متغیرهای مخفی و همچنین باورهای هرکدام از گره‌ها بوجود می‌آید. به عبارتی سایز پیام ها ارسالی و سایز باورهای هر متغیر در هر قاب با توجه به تعداد حالات هر فریم تغییر می‌یابد. ما با توجه به این مسئله تمهیداتی در نظر گرفتیم تا با توجه به قاب‌ها که دارای تعداد حالات مخفی متفاوت هستند، پیام های ارسال شده هم با توجه به آنها سایز مناسب خود را داشته باشد.

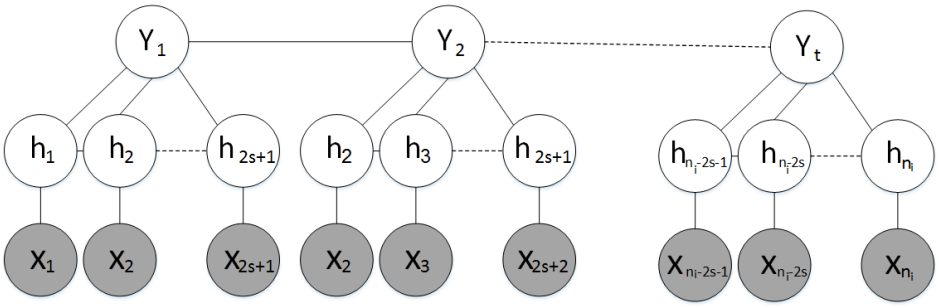
۲-۳-۴-مرحله آزمون

همچنین در مرحله آزمون با توجه به اتصال برچسب‌ها به هم استنتاج نسبت به مرحله آموزش متفاوت می‌شود. با توجه به مشاهده نشدن برچسب-های هر پنجره در این مرحله، فاکتور بین برچسب‌ها می‌تواند به عنوان پلی برای ارتباط بین متغیرهای مخفی لایه اول استفاده شود. این مدل گرافی یک شبه زنجیر است و می‌توانیم از الگوریتم انتشار باور درخت گروهک استفاده می‌کنیم به طوری که بتواند پیام‌های تولید شده را بین قسمت‌های مختلف منتشر کند. پیاده‌سازی برای انجام عمل استنتاج با درخت گروهک بر روی این مدل گرافی نیز انجام شده است. بنابراین برچسب‌های هر قاب را با استفاده از پارامترهای آموزش داده شده به صورت زیر محاسبه مینماییم.

$$y^* = \arg\max_y P(y|x,\theta^*) = \quad (8)$$
$$= \arg\max_y \sum_h P(y,h \mid x,\theta^*)$$

۵-آزمایشات

برای ارزیابی مدل، ما از پایگاه داده Arm Gesture استفاده می‌کنیم. این پایگاه داده یکی از مهمترین پایگاه داده‌های استفاده شده برای تشخیص حالت حرکت است. پایگاه داده Arm Gesture دارای شش برچسب است که هرکدام از برچسب‌ها مربوط به یک حالت خاص از حرکت‌های پشت سر هم دست است. ضربه زدن به عقب، حرکت عمودی جزئی، بازکردن به صورت عمودی، دو حرکت برگشت به عقب،اشاره کردن و برگشت به عقب و بازکردن به صورت افقی برچسب‌های این پایگاه داده می‌باشند[4]. در مرحله استخراج ویژگی، از هر قاب بیست ویژگی خام استخراج شده است از جمله مختصات اقلیدسی سه بعدی شانه‌ها و زاویه مشترک دوبعدی شانه های چپ و راست و آرنج. از سیزده نفر برای تهیه این پایگاه داده استفاده شده است و برای هر کلاس به طور میانگین 120 نمونه تهیه شده است.



شکل 4-مدل HCRF پنجره‌ای با جزئیات. وزن‌های مربوط به متغیرها در هر قاب غیرمشترک هستند و تعداد حالات در هر قاب پنجره می‌تواند متفاوت باشد و طول دنباله ثابت در نظر گرفته می‌شود.

برای مقایسه ما از مدل‌های مختلفی از قبیل HMM، CRF، HCRF استفاده کرده‌ایم. البته از انواع جدیدی که با توجه به مدل HCRF معرفی شده‌اند، شامل مدل‌های MM-HCRF، Linked HCRF، Coupled HCRF نیز برای مقایسه ذکر شده‌اند. همچنین از سایر روش‌های مهم استفاده شده در حوزه تشخیص حالت حرکت دست از قبیل روش سلسله مراتبی HSS و روش کاهش ابعاد ویژگی S-KDR-SVM استفاده شده است. جدول 1 دقت مدل‌های مختلف بر روی پایگاه داده Arm Gesture را نشان می‌دهد. در این پایگاه داده مدل‌های تفکیکی عملکرد بهتری نسبت به مدل‌های مولد داشته‌اند. همچنین در بین مدل‌های تفکیکی، مدل‌های دارای لایه مخفی عملکرد مناسب‌تری داشته‌اند. مدل سلسله مراتبی HSS در بین پژوهش‌های قبلی از عملکرد بهتری برخوردار است ولی مدل پیشنهادی دقت بیشتری درمقایسه با سایر مدل‌ها دارا می‌باشد.

جدول 1- دقت مدل‌های مختلف روی پایگاه داده Arm Gesture

Models	Mean Accuracy
HMM [4]	84.22
CRF [4]	86.03
HCRF [4]	91.64
HCRF (w=1) [4]	93.86
MM-HCRF [6]	93.79
S-KDR-SVM [6]	95.30
Linked HCRF [6]	97.65
Coupled HCRF [6]	97.24
HSS [11]	99.59
WHCRF	**99.84**

۶- نتیجه گیری و پیشنهادات

میدان تصادفی شرطی با حالات پنهان یکی از مدل‌های گرافی احتمالاتی تفکیکی با چارچوبی قدرتمند برای پردازش داده‌های دنباله‌ای به شمار می‌رود. ما در این مقاله مدل HCRF پنجره‌ای مبتنی بر پارامترهای غیرمشترک را معرفی کردیم. این مدل برای هر قاب موجود در پنجره تعریف شده وزن‌های متفاوتی را یاد گرفته و تعداد حالات مخفی آن نیز می‌توانند برای هر متغیر مخفی به صورتت دستی در مدل انتخاب شوند.

همچنین دقت مدل را بر روی پایگاه داده تشخیص حالت دست بررسی کرده و نشان دادیم با استفاده از پارامترهای غیرمشترک و تعداد حالات متفاوت و انجام عمل استنتاج با ارتباط دادن بین برچسب‌های هر پنجره می‌توان عملکرد بهتری را نسبت به سایر روش‌های گرافی احتمالاتی داشت.

مراجع

[1] A. Kapoor and R. Picard, "A R eal-Time Head Nod and Shake Detector,"Proc. Workshop Perspective User Interfaces, Nov. 2001.

[2] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, 2001

[3] C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas. Conditional models for contextual human motion recognition. In *ICCV*, 2005.

[4] A. Quattoni, S. B. Wang, L.-P. Morency, M. Collins, and T. Darrell. Hidden conditional random fields. *PAMI*, 29(10):1848–1852, 2007.

[5] L.-P. Morency, A. Quattoni, and T. Darrell. Latent-dynamic discriminative models for continuous gesture recognition. In *CVPR*, 2007.

[6] Yale Song, L. Morency, and R. Davis. "Multi-View Latent Variable Discriminative Models for Action Recognition," 2120–27, *CVPR.*,2012.

[7] A.K.H. Kuensch and S. Geman, "Hidden Markov Random Fields," Annals of Appied Probability, vol. 5, 2005.

[8] Y. Wang and G. Mori. Max-margin hidden conditional random fields for human action recognition. In *CVPR*, pages 872–879, 2009.

[9] Y. Wang and G. Mori. Hidden part models for human action recognition: Probabilistic versus max margin. *PAMI*, 33(7), 2011.

[10] D. Koller and N. Friedman, Probabilistic graphical models: principles and techniques. Cambridge, MA: *MIT Press*, 2009

[11] Yale Song, Louis-Philippe Morency, and Randall Davis. Action recognition by hierarchical sequence summarization. In *CVPR*, 2013.

[12] A. Shyr, R. Urtasun, and M. I. Jordan. Sufficient dimension reduction for visual sequence classification. In *CVPR*, 2010.
[13] D. Yu and L. Deng. Deep-structured hidden conditional random fields for phonetic recognition. In *INTERSPEECH*, 2010
[14] D. Yu, S. Wang, Z. Karam, and L. Deng, "Language recognition using deep-structured conditional random fields," 2010, pp. 5030–5033.
[15] J. Peng, L. Bo, and J. Xu. Conditional neural fields. In *NIPS*, 2009
[16] C. A. Sutton, K. Rohanimanesh, and A. McCallum. Dynamic conditional random fields: factorized probabilistic models for labeling and segmenting sequence data. In *ICML*, 2004.
[17] C. Sutton, K. Rohanimanesh, and A. McCallum. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. In *JMLR,2007*.
[18] N. Chen, J. Zhu, and E. Xing. Predictive subspace learning for multi-view data: a large margin approach. In *NIPS*, pages 361–369, 2010.
[19] Y. Fujii, K. Yamamoto, and S. Nakagawa, "Automatic speech recognition using Hidden Conditional Neural Fields," 2011, pp. 5036–5039.
[20] Yu, D., Deng, L., and Wang, S., "Learning in the deep structured conditional random fields," Proc. *NIPS Workshop*, Dec. 2009
[21] L. Ladicky, C. Russell, P. Kohli, P. H. S. Torr, "Associative hierarchical CRFs for object class image segmentation", in Proc. *ICCV* 2009.
[22] L. Liao, D. Fox, and H. Kautz, "Hierarchical conditional random fields for GPS-based activity recognition", in Proc. of International Symposium of Robotis Research (*ISRR*), 2007
[23] Liu, C., Liu, J., He, Z., Zhai, Y., Hu, Q., & Huang, Y. Convolutional neural random fields for action recognition. *Pattern Recognition*. 2016.
[24] Yale Song and R. Davis. "Continuous Body and Hand Gesture Recognition for Natural Human-Computer Interaction Extended Abstract, *IJCAI*.,2015.