



دانشگاه سمنان

دانشکده مهندسی برق و کامپیوتر

طرح پژوهشی

تحلیل تصویر با استفاده از میدان تصادفی شرطی

مجری:

دکتر محمدجواد فدائی اسلام

مهر ۱۳۹۹

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فهرست مطالب

۱	فصل اول: مقدمه
۱	۱-۱- تعاریف اولیه
۷	۲-۱- هدف طرح
۸	۳-۱- ساختار طرح پژوهشی
۹	فصل دوم: میدان تصادفی شرطی
۱۰	۱-۲- مدل مارکوف مخفی
۱۳	۲-۲- میدان تصادفی شرطی
۱۵	۳-۲- چند مثال از گراف بدون جهت
۱۷	۴-۲- میدان تصادفی شرطی زنجیره‌ای خطی
۲۳	فصل سوم: کاربردهای میدان تصادفی شرطی
۲۴	۱-۲- کاربردهای میدان تصادفی شرطی
۲۴	۳-۲- میدان تصادفی شرطی در قطعه‌بندی معنایی تصویر
۳۰	۳-۳- جمع‌بندی کاربرد میدان تصادفی در حوزه قطعه‌بندی
۳۲	فصل چهارم: توسعه کاربرد میدان تصادفی شرطی در استخراج نواحی برجسته تصویر
۳۳	۱-۴- مقدمه
۳۵	۲-۴- تشخیص برجستگی بر اساس محتوا
۳۹	۳-۴- معیارهای ارزیابی
۴۲	۴-۴- پایگاه داده‌ها
۴۳	۵-۴- روش پیشنهادی
۴۷	۴-۶- پیاده‌سازی و نتایج
۵۵	۷-۴- نتیجه‌گیری
۵۶	فصل پنجم جمع‌بندی و کارهای آینده
۵۷	۱-۵- جمع‌بندی
۵۸	۲-۵- کارهای آینده
۵۹	مراجع

فصل اول

مقدمه

پیش‌بینی چندین متغیر به هم وابسته و کشف ارتباط بین آنها هدف بسیاری از کاربردهاست. چنین کاربردهایی بسیار متنوع هستند، بخش‌بندی نواحی تصویر، تخمین امتیاز دهی در بازی Go، بخش‌بندی ژن‌ها در رشته DNA و پارس نحوی متون زبان طبیعی، نمونه‌هایی از این کاربردها را نشان می‌دهند [۱]. در چنین کاربردهایی ما علاقه داریم یک بردار خروجی از متغیرهای تصادفی $\mathbf{y} = \{y_0, y_1, \dots, y_T\}$ را از روی یک بردار ویژگی \mathbf{x} پیش‌بینی کنیم. مدل‌های گرافیکی^۱ راهی مناسب برای نمایش وابستگی بین متغیرهاست. با مدل‌های گرافیکی می‌توان یک توزیع پیچیده بر روی تعداد زیادی از متغیرها را به صورت ضربی از فاکتورهای محلی با تعداد کمتری از متغیرها نمایش داد. میدان تصادفی شرطی یکی از این مدل‌های موفق است. در این فصل ابتدا تعاریف اولیه مطرح خواهند شد تا جایگاه روش‌های مختلف یادگیری باز تعریف شود.

۱-۱- تعاریف اولیه

از دیدگاه روش‌های یادگیری با ناظر مدل‌های آماری را می‌توان به دو دسته مدل‌های مولد^۲ و مدل‌های تفکیکی^۳ تقسیم کرد. در ادامه به بررسی این تقسیم‌بندی می‌پردازیم.

۱-۱-۱- مدل تفکیکی و مولد

در مدل‌های تفکیکی مانند رگرسیون، ماشین دسته‌بند بردار پشتیبان^۴ شبکه عصبی چند لایه^۵ برای پیش‌بینی تعلق داده ورودی به کلاس y از روی داده‌های آموزش \mathbf{x} ، محاسبه زیر باید انجام شود:

$$f(\mathbf{x}) = \arg \max_y p(y|\mathbf{x}) \quad (1-1)$$

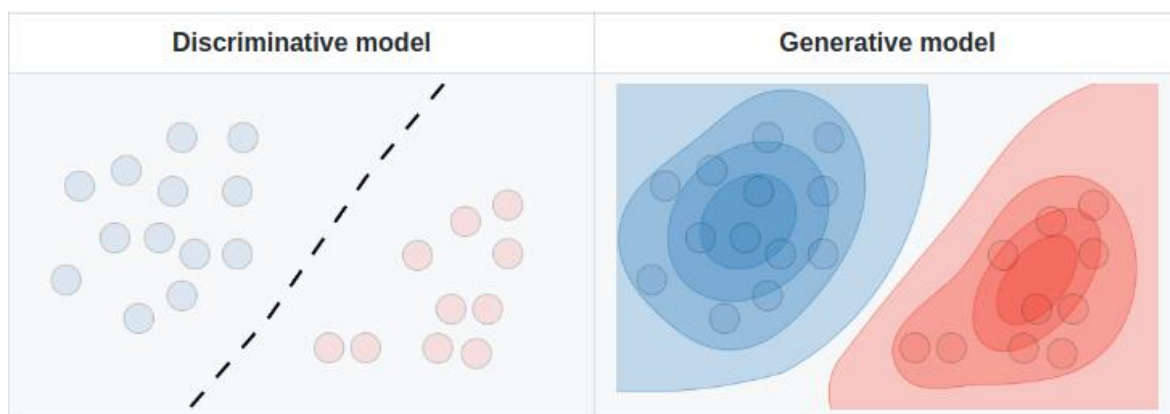
محاسبه بالا مشابه آن است که یک مرز جداکننده بین دو کلاس رسم شود (شکل (۱-۱)). با استفاده از قانون بیز می‌توان $p(y|\mathbf{x})$ را به صورت زیر نوشت:

$$p(y|\mathbf{x}) = \frac{p(y)p(\mathbf{x}|y)}{p(\mathbf{x})} \quad (2-1)$$

از آنجا که مخرج برای همه y ها یکسان است؛ می‌توان آن را نادید گرفت و رابطه (۱-۱) را به صورت زیر در نظر گرفت:

$$f(x) = \arg \max_y p(y)p(x|y) \quad (3-1)$$

این رابطه‌ای است که در مدل‌های مولد استفاده می‌شود. در رابطه (۱-۱) یک توزیع احتمال شرطی $p(y|x)$ محاسبه می‌شود که مرز بین کلاس‌ها را مدل می‌کند. اما در رابطه (۳-۱) باید توزیع احتمال توام $p(x,y)$ محاسبه شود؛ زیرا $p(x,y) = p(y)p(x|y)$. در این حالت توزیع هر کلاس باید مدل شود. شکل (۱-۱) عمل کرد این دو نگرش نشان داده شده است.



شکل (۱-۱) نمایش کارکرد مدل مولد و تفکیکی

همان‌طور که مطرح شد، مدل‌های مولد احتمال شرطی حالات ارائه شده توسط مشاهدات را با استفاده از احتمال توام به دست می‌آورند؛ اما محاسبه دقیق مدل‌های احتمالی معمولاً غیر ممکن می‌باشد، بنابراین فرض استقلال شرطی در نظر گرفته می‌شود تا مدل احتمالی قابل حل باشد [۲]. فرض استقلال شرطی، فرض می‌کند که هر مشاهده با توجه به حالت آن مستقل از مشاهدات دیگر است در این صورت داریم:

$$p(x|y) = \prod_{i=1}^N p(x_i|y) \quad (4-1)$$

مدل بیز ساده^۱ و مدل مارکوف مخفی^۲ از مدل‌های مولد به شمار می‌روند. مدل بیز ساده از جمله روش‌هایی است که از فرض استقلال برای سادگی بهره می‌برد و به صورت زیر است:

$$p(x,y) = p(y)p(x|y) = p(y) \prod_{i=1}^N p(x_i|y) \quad (5-1)$$

داده‌های عادی معمولاً ابعادی بالا و وابسته به هم دارند، بنابراین فرض استقلال شرطی روی x فرض

۱ - naïve bayes

۲ - hidden markov model-HMM

مناسبی نیست. بنابراین دقت مدل‌سازی مدل‌های مولد که از این فرض بهره می‌برند نسبت به مدل‌های تفکیکی کمتر است، زیرا مدل‌های تفکیکی، مدل‌سازی مستقیم احتمال شرطی را فراهم می‌کنند [۳]. اگرچه مدل‌های مولد در مقایسه با رویکردهای تفکیکی در دقت مدل‌سازی دارای اشکالاتی هستند، اما دارای مزایایی نیز هستند:

- مدل مولد می‌تواند دارای مقادیر مفقود شده باشد، زیرا داده‌ها برای هر کلاس به صورت جداگانه مدل‌سازی می‌شوند. این مدل می‌تواند مجموعه کوچکی از داده‌های برچسب خورده را با مجموعه بزرگی از داده‌های بدون برچسب افزایش دهد، زمانی که داده‌های برچسب‌دار گران هستند.
- یک کلاس جدید از داده‌ها را به راحتی می‌توان به کل مسئله طبقه‌بندی توسعه داد، زیرا داده‌ها به کلاس مدل‌سازی شده وابسته هستند [۴].

در مدل بی‌ساده، تنها یک متغیر خروجی در نظر گرفته می‌شود. اما در برخی مسائل، گاهی هدف پیش‌بینی دنباله یا برداری از کلاس‌ها برای دنباله‌ای از مشاهده‌ها است. مدل‌های مارکوف مخفی برای حل این گونه مسائل ارائه شده‌اند. در مدل مارکوف مخفی دنباله‌ای از مشاهده‌ها را خواهیم داشت که دنباله‌ای از برچسب‌ها (یکی برای هر مشاهده) را باید برای آن پیش‌بینی کرد. بنابراین می‌توان مدل مارکوف مخفی را دنباله‌ای از بی‌ساده‌های ساده در نظر گرفت که به هم متصل شده‌اند. در پیاده‌سازی مدل مارکوف مخفی دو فرض در نظر گرفته شده است. اول: فرآیند مارکوفی است یعنی هر وضعیت تنها به وضعیت قبل خود وابسته است. به عبارتی y_i تنها به y_{i-1} وابسته است و از $y_1, y_2, y_3 \dots y_{i-2}$ مستقل است. دوم: توجه داشته باشید که تنها یک مشاهده یا ویژگی در هر وضعیت دنباله رخ می‌دهد، یا به عبارتی در وضعیت y_i تنها مشاهده x_i رخ می‌دهد. از این رو در مدل مارکوف مخفی داریم:

$$p(\mathbf{x}, \mathbf{y}) = \prod_{t=1}^T p(y_t | y_{t-1}) p(x_t | y_t) \quad (6-1)$$

وابستگی یک مشاهده به تنها یک وضعیت از نواقص مدل مخفی مارکوف است. میدان‌های تصادفی شرطی که بعداً به آن خواهیم پرداخت دقیقاً برای رفع این مشکل ارائه شده‌اند [۲][۱]. همان‌طور که مطرح شد رویکردهای تفکیکی، یک مسئله را به صورت احتمال شرطی از حالت‌های \mathbf{y} یا متغیرهای درونی با توجه به مشاهدات \mathbf{x} ، مستقیماً مدل می‌کنند. بنابراین، عملکرد کلی مدل‌های تفکیکی به دلیل تفاوت بین مدل و توزیع حقیقی داده‌ها، بیشتر از مدل‌های مولد می‌باشد [۳]. دو مزیت مدل‌های تفکیکی عبارتند از:

- مدل‌های تفکیکی سریع‌تر از مدل‌های مولد هستند، زیرا آن‌ها به جای یک روش تکراری، به طور مستقیم نقطه داده‌های جدید را پیش‌بینی می‌کنند تا احتمال نقطه داده در هر کلاس پیدا شود.

- انتظار می‌رود که روش‌های تفکیکی عمل‌کرد پیش‌بینی بهتری نسبت به روش‌های مولد داشته باشند، زیرا مدل‌های تفکیکی برای پیش‌بینی برچسب کلاس به جای محاسبه توزیع توام بردارهای ورودی و اهداف، آموزش دیده‌اند [۵].

۱-۲-۱- نمایش گرافیکی

مدل گراف احتمالی، نمایش یک توزیع احتمال به صورت یک گراف است. در چنین گرافی، به ازای هر یک از متغیرهای تصادفی گره‌ای وجود دارد. عدم وجود یال یا مسیر بین دو متغیر، نشان دهنده استقلال آنها است. استقلال شرطی به این معنی است که دو متغیر تصادفی a و b با توجه به متغیر سوم c مستقل هستند، اگر $p(a, b|c) = p(a|c)p(b|c)$. از چنین گراف‌هایی که گراف‌های استقلال نیز نامیده می‌شوند، می‌توان خواص استقلال شرطی توزیع پایه^۲ را درک کرد. توجه داشته باشید که یک گراف کاملاً متصل اطلاعات چندانی درباره ارتباط متغیرها نمی‌دهد، این عدم وجود یال‌هاست که حاوی اطلاعات مفیدی است. استقلال شرطی، مفهوم مهمی است که می‌توان از آن برای تجزیه‌ی توزیع‌های احتمال پیچیده به ضرب یک سری از عامل‌ها استفاده کرد، هر یک از این عامل‌ها شامل زیر مجموعه‌ای از متغیرهای تصادفی متناظر آن است. این مفهوم، محاسبات پیچیده به خصوص در فازهای یادگیری و استنتاج را بسیار کاهش می‌دهد. به طور کلی، تجزیه یا در واقع عامل‌بندی یک توزیع احتمال، به صورت ضرب عامل‌ها با علامت ψ_s نشان داده می‌شود که در آن v_s زیر مجموعه‌ای از متغیرهای تصادفی تشکیل دهنده‌ی عامل است.

$$P(\mathbf{v}) = \prod_s \psi(v_s) \quad (7-1)$$

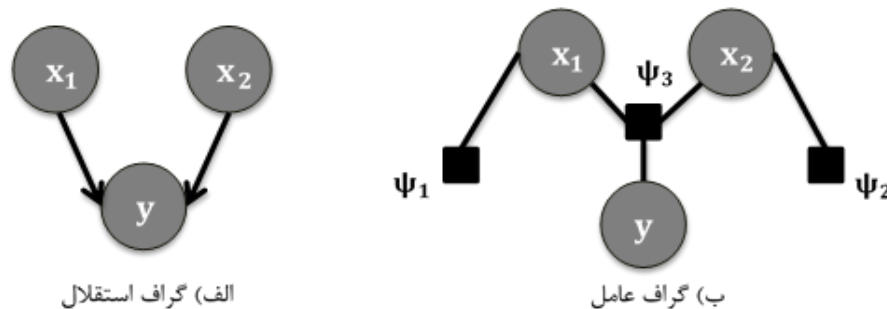
یک گراف استقلال با $\mathbf{G}=(\mathbf{V},\mathbf{E})$ نشان داده می‌شود که \mathbf{V} در آن نشان‌دهنده راس و \mathbf{E} نشان‌دهنده یال است. نمونه‌ای از آن در شکل (۱-۲ الف) نشان داده است. راس‌ها مجموعه‌ای از متغیرهای تصادفی هستند که با دایره نشان داده می‌شوند و از دو دسته X و Y تشکیل شده‌اند. معمولاً X به عنوان مجموعه‌ی متغیرهای ورودی یا مشاهده، و Y به عنوان مجموعه‌ی متغیرهای خروجی در نظر گرفته می‌شود. گراف استقلال با توجه به نوع مدل گرافیکی که نشان‌دهنده‌ی آن است می‌تواند به دو صورت جهت‌دار و یا بدون جهت باشد.

۱- independency graphs

۲- underlying distribution

در یک گراف عامل، همانند گراف استقلال، متغیرهای تصادفی مربوط به توزیع پایه به وسیله‌ی دایره نشان داده می‌شوند (شکل ۲-۱ ب). علاوه بر این، گراف‌های عامل شامل گره‌های عامل نیز می‌باشند که به وسیله مربع‌های کوچک پر رنگ، که نشان دهنده عامل‌های ψ_s می‌باشند، به تصویر کشیده می‌شوند. در یک گراف عامل، یال‌ها همواره بدون جهت هستند و متغیرهای تصادفی را به گره‌های عامل متصل می‌کنند. عامل ψ_s شامل تمام متغیرهای تصادفی می‌باشد که گره عامل مربوط به آن‌ها، به طور مستقیم توسط یک یال متصل شده باشد. بنابراین، گراف عامل به صراحت نشان دهنده‌ی عامل‌بندی توزیع احتمال پایه می‌باشد. گراف استقلال هر دو مدل گرافیکی جهت‌دار و یا بدون جهت را می‌توان به یک گراف عامل تبدیل نمود.

به‌عنوان مثال توزیع احتمال $P(x_1, x_2, y)$ به صورت $P(x_1)P(x_2)P(y|x_1, x_2)$ عامل‌بندی می‌شود. در این عامل‌بندی، عامل‌های $\psi_1(x_1) = P(x_1)$ ، $\psi_2(x_2) = P(x_2)$ و $\psi_3(y, x_1, x_2) = P(y|x_1, x_2)$ وجود دارند. در اینجا x_1 و x_2 با توجه به y مستقل شرطی هستند. شکل (۲-۱) نشان دهنده‌ی گراف استقلال و گراف عامل این توزیع است [۲]. در ادامه مدل‌های گرافیکی جهت‌دار و بدون جهت، بررسی خواهند شد.



شکل (۲-۱) مدل گرافیکی جهت‌دار [۲]

مدل‌های گرافیکی جهت‌دار

مدل‌های گرافیکی جهت‌دار، خانواده‌ای از توزیع‌ها هستند که به صورت زیر عامل‌بندی می‌شوند:

$$p(\mathbf{y}) = \prod_{s=1}^S p(y_s | \mathbf{y}_{\pi(s)}) \quad (۸-۱)$$

در این عامل‌بندی $\mathbf{y}_{\pi(s)}$ نشان‌دهنده والد‌های گره y_s است. مجموعه $\pi(s)$ می‌تواند تهی باشد که در این صورت $p(y_s | \mathbf{y}_{\pi(s)})$ به $p(y_s)$ تبدیل می‌شود [۱]. طبقه‌بندی‌کننده بیز ساده را در نظر بگیرید. شکل (۳-۱) نمایش گرافیکی چنین مدلی را برای سه متغیر مشاهده شده، نشان می‌دهد. توزیع احتمال متناظر

با فرض بیز ساده را می توان به صورت زیر عامل بندی کرد.

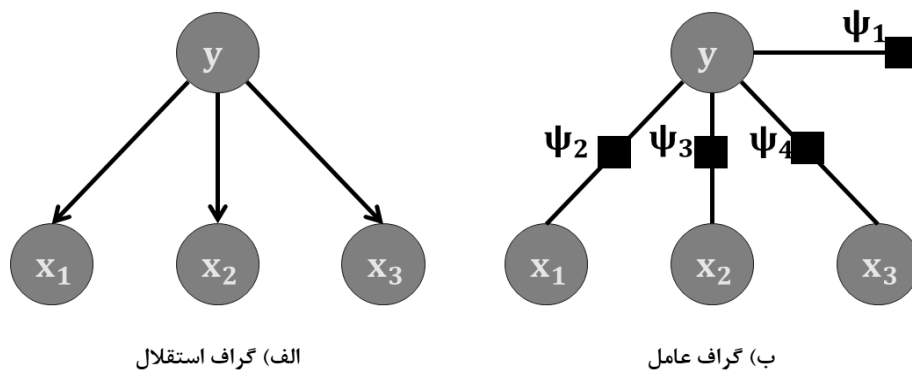
$$p(y, x_1, x_2, x_3) = p(y) \cdot p(x_1|y) \cdot p(x_2|y) \cdot p(x_3|y) \quad (9-1)$$

به طور مشابه، شکل (۴-۱) یک طبقه بندی کننده مدل مارکوف مخفی را برای دنباله ای با سه متغیر ورودی

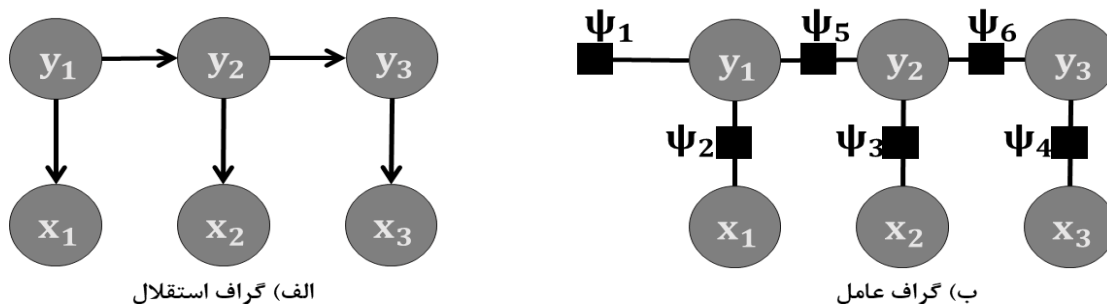
x_1, x_2 و x_3 نشان می دهد. عامل بندی متناظر با آن براساس رابطه (۸-۱) به صورت زیر است:

$$p(x_1, x_2, x_3, y_1, y_2, y_3) = \psi_1(y_1) \cdot \psi_2(x_1, y_1) \cdot \psi_3(x_2, y_2) \cdot \psi_4(x_3, y_3) \cdot \psi_5(y_1, y_2) \cdot \psi_6(y_2, y_3) \quad (10-1)$$

معمولا مدل گرافیکی جهت دار را برای مدل های مولد نظیر بیز ساده یا مدل مارکوف مخفی به کار می برند.



شکل (۳-۱) طبقه بندی ساده به عنوان یک گراف جهت دار و گراف عامل متناظر با آن [۲]



شکل (۴-۱) گراف استقلال و گراف عامل برای مدل مارکوف مخفی [۲]

مدل های گرافیکی بدون جهت

یک توزیع احتمال را می توان با استفاده از حاصل ضرب توابع غیر منفی کلیک های حداکثر گراف G ، به صورت یک مدل گرافیکی بدون جهت نمایش داد. این عامل بندی به گونه ای انجام می شود که گره های مستقل شرطی درون عامل یکسان ظاهر نمی شوند، این بدان معنی است که آن ها به کلیک های مختلف تعلق دارند:

$$P(y) = \frac{1}{Z} \prod_{c \in C} \psi_c(y_c) \quad (11-1)$$

عامل‌های $\psi_c \geq 0$ که توابع پتانسیل نیز نامیده می‌شوند، از متغیرهای تصادفی \mathbf{y}_c درون یک کلیک $c \in C$ هستند. توابع پتانسیل می‌توانند به صورت هر تابع دلخواهی باشند. با توجه به این اصل کلی، توابع پتانسیل لزوماً نباید توابع احتمالی باشند. این نکته در تضاد با گراف‌های جهت‌دار می‌باشد، زیرا گراف‌های جهت‌دار توزیع توام را به صورت حاصل ضرب توزیع‌های شرطی عامل‌بندی می‌کنند. بنابراین، نرمال‌سازی حاصل ضرب توابع پتانسیل برای دستیابی به یک احتمال مناسب ضروری است. این امر توسط عامل نرمال‌ساز Z انجام می‌شود. محاسبه Z یکی از چالش‌های مهم در حین یادگیری پارامتر می‌باشد، زیرا برای محاسبه آن نیاز به جمع همه‌ی انتساب‌های ممکن برای \mathbf{y} است:

$$Z = \sum_{\mathbf{y}} \prod_{c \in C} \psi_c(\mathbf{y}_c) \quad (12-1)$$

مدل‌های گرافیکی جهت‌دار و بدون جهت، در روش توزیع احتمال اصلی متفاوت هستند. عامل‌بندی در مدل‌های گرافیکی جهت‌دار مستقیماً از طریق حاصل ضرب توزیع‌های احتمال شرطی انجام می‌گیرد. در مدل‌های گرافیکی بدون جهت، عامل‌بندی به وسیله توابع دلخواه انجام می‌شود. در این روش نیازی به ذکر صریح اینکه چگونه این متغیرها به هم مرتبط هستند، نمی‌باشد. اما این روش به دلیل نیاز به محاسبه عامل نرمال‌ساز هزینه‌ی بیشتری را دارا می‌باشد [۲].

یک مدل مولد را می‌توان به صورت زیر نیز در قاعده بیز قرار داد:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y}|\mathbf{x}) \quad (13-1)$$

در این حالت مشاهده مقدم بر خروجی آمده است؛ برای آنکه این ترتیب حفظ شود مدل گرافیک جهت‌دار را برای روش‌های مولد در نظر می‌گیرند و مدل بدون جهت را برای روش‌های تفکیکی. اما این عمومیت ندارد و مدل‌های بدون جهت هم برای روش‌های مولد و مدل جهت‌دار برای روش‌های تفکیکی به کار رفته است [۲].

۱-۲- هدف طرح

علاقه زیادی برای به کارگیری میدان تصادفی شرطی در مسائل گوناگون وجود دارد. کاربردهای موفقی از آن در پردازش متن، بیوانفورماتیک و بینایی ماشین دیده شده است. اگرچه کاربردهای اولیه میدان تصادفی شرطی به صورت زنجیره خطی بود، اما اخیراً کاربردهایی با ساختار گرافی متنوع‌تری از آن مشاهده شده است. این نوع از میدان‌های تصادفی شرطی در مسایل پیچیده‌تری مثل گراف‌ها و درخت‌ها می‌تواند کاربرد داشته باشد. هدف این طرح، در گام اول، شناسایی دقیق‌تر این مدل و آشنایی با کاربردهای آن در حوزه پردازش تصویر است. در گام بعدی این مدل را برای یکی از نیازهای پردازش تصویر که شناسایی نواحی برجسته تصویر است پیاده‌سازی نمودیم. شناسایی اهمیت یک ناحیه از تصویر زمانی کارا خواهد بود که از

اطلاعات نواحی همسایه نیز استفاده شود. میدان تصادفی شرطی یکی از مدل‌های موفق در کشف ارتباط بین نواحی است. از این رو به نظر رسید به کارگیری آن در این کاربرد بتواند بهبود زیادی در نتایج ایجاد نماید.

۱-۳- ساختار طرح پژوهشی

پس از بررسی تعاریف اولیه و جایگاه میدان تصادفی شرطی در حوزه هوش مصنوعی که در مقدمه مطرح شدند. مدل‌سازی، استنتاج و تخمین پارامتر در میدان تصادفی شرطی در فصل دوم بررسی شده است. کاربردهای میدان تصادفی شرطی در پردازش تصویر در فصل سوم بررسی شده است. توسعه میدان تصادفی شرطی در حوزه شناخت برجستگی تصویر و آرایه مدلی بر اساس آن و نتایج پیاده‌سازی در فصل چهار آمده است. جمع‌بندی و پیشنهاد برای کارهای آتی عنوان فصل پایانی این طرح است.

فصل دوم

میدان تصادفی شرطی

در فصل قبل، برخی مفاهیم اولیه در حوزه یادگیری یادآوری شد. دو رهیافت عمده در روش‌های یادگیرنده که عبارت بودند از روش‌های مولد و تفکیکی مورد بررسی قرار گرفت و نقاط قوت و ضعف آنها مطرح گردید. از سوی دیگر نمایش گرافیکی مدل‌های احتمالی به عنوان روشی موثر در تحلیل و ساده‌سازی روش‌های آماری بیان گردید. به طور ضمنی جایگاه روش‌های میدان تصادفی شرطی به عنوان مدل یادگیرنده تفکیکی با نمایش گرافیکی بدون جهت مشخص شد. در این فصل به صورت دقیق‌تر میدان تصادفی شرطی را بررسی خواهیم نمود. مدل مارکوف مخفی از روش‌هایی است که بر روی دنباله‌ها کار می‌کند. برای بررسی بهتر میدان تصادفی شرطی روش مدل مارکوف مخفی بررسی خواهد شد و تمایز آن با میدان تصادفی شرطی بیان می‌شود. آموزش و استنتاج از مراحل مهم در پیاده‌سازی یک میدان تصادفی شرطی است که با استفاده از مثال به شرح آن خواهیم پرداخت.

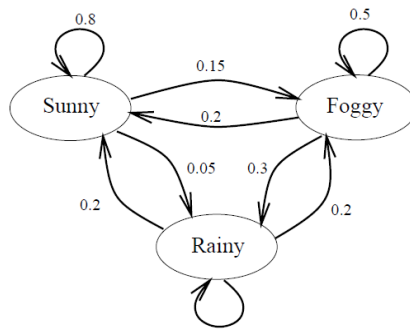
۲-۱- مدل مارکوف مخفی

ابتدا مدل مارکوف را تعریف می‌کنیم. اگر یک دنباله در وضعیت n تنها به وضعیت آن در $n - 1$ وابسته باشد؛ می‌گوییم این دنباله از نوع مدل مارکوف مرتبه اول است. به طور مثال اگر وضعیت هوا در روز n را با نماد ω_n نشان دهیم، با فرض مارکوف مرتبه اول می‌توان به صورت زیر نوشت:

$$P(\omega_n | \omega_{n-1} \omega_{n-2} \omega_{n-3} \dots \omega_1) = P(\omega_n | \omega_{n-1}) \quad (1-2)$$

رابطه بالا بیان می‌دارد که وضعیت هوا فقط به روز قبل وابسته است و به روزهای قبل‌تر وابسته نیست. اگر دنباله در وضعیت n به دو حالت قبل $n - 1$ و $n - 2$ وابسته باشد مارکوف مرتبه دوم نام دارد. در اصطلاح، مدل مارکوفی به مدلی گفته می‌شود که فرض مارکوف مرتبه اول برای آن صادق باشد. اگر وضعیت یک پدیده برای ما معین نباشد و تنها مشاهداتی داشته باشیم که از وضعیت پدیده حکایت کند مساله ما از جنس مدل مارکوف مخفی است. به طور مثال اگر در مساله هواشناسی نوع آب و هوا برای ما قابل رویت نباشد (در زندان باشیم) و تنها مشاهده‌ای مثل در دست دادن چتر توسط نگهبان در دسترس باشد مساله ما از جنس مدل مارکوف مخفی است. در بسیاری از پدیده‌های اطراف ما این حالت وجود دارد. یعنی مشاهداتی وجود دارد که باید پی به حالت‌ها برده شود.

سه نوع اطلاعات یا به عبارتی سه ماتریس برای تشکیل مدل مارکوف مخفی باید از داده‌ای موجود استخراج شود. ماتریس فضای حالت، ماتریس احتمال مشاهدات در هر حالت و ماتریس شرایط اولیه که احتمال بودن در هر یک از حالات در روز اول را نشان می‌دهد. این سه نوع اطلاعات در مثال هواشناسی در شکل (۱-۲) و شکل (۲-۲) نشان داده شده است. ماتریس فضای حالت را با A نشان می‌دهند و ماتریس احتمال رخداد مشاهدات را با B و ماتریس حالت اولیه را با π نشان می‌دهند.



شکل (۱-۲) نمای یک گراف فضای حالت در مساله آب و هوا

احتمال وضعیت اولیه (π)	احتمال مشاهده (داشتن چتر) (B)	وضعیت
۰,۵	۰,۱	Sunny
۰,۲	۰,۸	Rainy
۰,۳	۰,۳	Foggy

شکل (۲-۲) ماتریس احتمال مشاهده (داشتن چتر) در هر حالت و ماتریس احتمال وضعیت اولیه

فرمول بندی مدل مارکوف مخفی برای حل $p(\mathbf{x}, \mathbf{y})$ به صورت زیر خواهد بود:

$$p(\mathbf{x}, \mathbf{y}) = \prod_{t=1}^T p(y_t | y_{t-1}) p(x_t | y_t) \quad (1-2)$$

رابطه بالا دنباله حالتها (وضعیتها) را با \mathbf{y} و مشاهدات را با \mathbf{x} نشان داده شده است. این رابطه یک مدل مارکوف خطی را نمایش می‌دهد [۱]. برای ساخت مدل مارکوف مخفی سه ماتریس A ، B و π باید محاسبه شود

مدل مارکوف مخفی را می‌توان به صورت گراف عامل نشان داد:

$$p(\mathbf{x}, \mathbf{y}) = \prod_t \Psi_t(y_t, y_{t-1}, x_t) \quad (2-2)$$

در رابطه بالا فاکتورها را می‌توان به صورت زیر نشان داد:

$$\Psi_t(j, i, x) \stackrel{\text{def}}{=} p(y_t = j | y_{t-1} = i) p(x_t = x | y_t = j) \quad (3-2)$$

در رابطه بالا $Z = 1$ است. اگر به صورت مدل گرافی در نظر بگیریم $\Psi_t(j, i, x)$ وزن انتقال از i به j وقتی که مشاهده جاری x است را نشان می‌دهد [۱]. سه مساله در مدل مارکوف مخفی طرح می‌گردد که از دیدگاه گراف عامل به بررسی آنها می‌پردازیم:

۱- یافتن احتمال رخداد یک مشاهده $p(x)$

برای محاسبه آن از فرمول بندی زیر کمک می‌گیریم

$$p(x) = \sum_y p(x, y) \quad (4-2)$$

در رابطه بالا اندیس Σ بردار y است که انواع دنباله‌های ممکن حالات را نشان می‌دهد. محاسبه مشاهده x بدون استفاده از تکنیک خاص دارای پیچیدگی محاسباتی نمایی است. برای حل آن روش پیشرو ابداع شده است.

۲- یافتن محتمل ترین رشته فضای حالت با توجه به مشاهده خاص x یا به عبارتی:

$$y^* = \arg \max_y p(y|x) \quad (5-2)$$

در این حالت مشاهده‌ای رخ داده است. رشته‌ای از حالات درونی را جستجو می‌کنیم که بیشترین احتمال رخداد x را ایجاد نماید. این مساله با استفاده از الگوریتم برنامه نویسی پویای ویتربی^۲ حل می‌شود.

۳- محاسبه احتمال $p(y_{t-1}, y_t | x)$ یا به عبارتی ساخت مدل احتمالی مارکوف با تولید سه ماتریس A ، B و π .

برای محاسبه مدل مارکوف باید از الگوریتم پس‌رو-پیش‌رو^۳ استفاده نمود. جزییات آن در [۶] و [۱] آمده است.

نقایص مدل مارکوف مخفی

علی‌رغم موفقیت چشم‌گیر مدل مارکوف مخفی برخی از نقایص آن در این زیر بخش ذکر شده است.

۱- در مدل مارکوف مخفی هر مشاهده به یک وضعیت وابسته است اما در میدان تصادفی شرطی هر

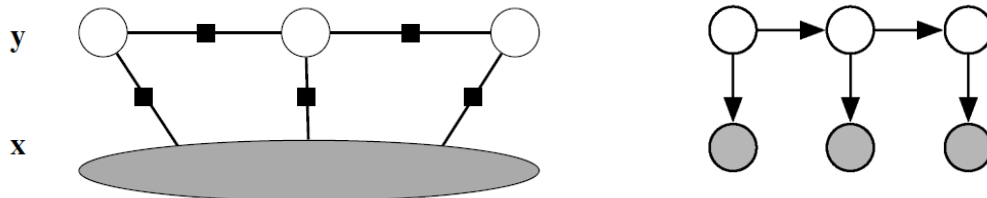
forward algorithm - ۱

Viterbi algorithm - ۲

forward-backward algorithm - ۳

مشاهده به تمام حالات وابسته است. این موضوع در شکل (۳-۲) نشان داده شده است. این موضوع به قابلیت‌های میدان تصادفی شرطی می‌افزاید.

۲- از آنجا که مدل مارکوف مخفی روشی مولد است، برای محاسبه آن نیازمند به محاسبه $p(\mathbf{x})$ هستیم، که همان‌طور که در فصل قبل مطرح شد؛ محاسبه آن سنگین است.



شکل (۳-۲) سمت راست: مدل مارکوف مخفی که هر مشاهده تنها به یک وضعیت وابسته است. سمت چپ: مدل میدان تصادفی شرطی که تمام مشاهدات به تمام وضعیت‌ها وابسته است.

۳- در ساخت مدل مارکوف تنها یک ویژگی می‌تواند بررسی شود. اضافه نمودن ویژگی‌های بیشتر محاسبات آن را پیچیده می‌نماید. اما در میدان تصادفی شرطی ویژگی‌های متعددی را می‌توان به صورت همزمان مورد استفاده قرار داد.

۲-۲- میدان تصادفی شرطی

میدان‌های تصادفی شرطی در سال ۲۰۰۱ توسط Lafferty و همکارانش ارائه شد، میدان‌های تصادفی شرطی مدل‌های احتمالی برای برچسب‌زنی داده‌های متوالی هستند [۷]. در واقع میدان‌های تصادفی شرطی مدلی احتمالی برای محاسبه‌ی احتمال رخداد یک زنجیره از حالت‌ها با توجه به زنجیره مشاهدات است. اگر G گراف عامل بر روی $F = \{\Psi_a\}$ مجموعه‌ای از فاکتورها در G باشد؛ آنگاه یک توزیع شرطی برای یک میدان تصادفی شرطی به صورت زیر است:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{a=1}^A \Psi_a(\mathbf{y}_a, \mathbf{x}_a) \quad (۶-۲)$$

در رابطه بالا تعداد فاکتورها A است که هر کدام با اندیس a مشخص می‌شوند. $\mathbf{y}_a, \mathbf{x}_a$ نشان‌دهنده مجموعه‌ای از حالت‌ها و مشاهدات است که در فاکتور Ψ_a وجود دارند. Z ثابت نرمال‌سازی نام دارد که از رابطه زیر محاسبه می‌شود:

$$Z(x) = \sum_y \prod_{a=1}^A \Psi_a(y_a \cdot x_a) \quad (7-2)$$

همان طور که در نقایص مدل مارکوف مخفی مطرح شد، از مزایای میدان تصادفی شرطی آن است که از چند ویژگی می‌توان استفاده نمود. از سوی دیگر با الهام از مدل‌های رگرسیون لجیستیک^۱ بهتر است $\log \Psi_a$ در برابر مجموعه از پیش تعیین شده ویژگی‌ها رفتار خطی داشته باشد. با در نظر گرفتن این دو موضوع Ψ به صورت زیر محاسبه می‌شود:

$$\Psi_a(y_a \cdot x_a) = \exp \left\{ \sum_{k=1}^{K(A)} \theta_{ak} f_{ak}(y_a \cdot x_a) \right\} \quad (8-2)$$

پارامتر θ_{ak} نشان‌دهنده وزن هر فاکتور است و باید در فرآیند آموزش محاسبه شود.

مدل مارکوف مخفی را می‌توان به عنوان نسخه‌ی دنباله‌ای از یک مدل بیز ساده در نظر گرفت: به جای یک تصمیم‌گیری مستقل، مدل مارکوف مخفی یک توالی خطی از تصمیمات را مدل می‌کند. بر این اساس، میدان تصادفی شرطی را نیز می‌توان، به عنوان نسخه دنباله‌ای (توالی) مدل‌های حداکثر آنتروپی (رگرسیون لجیستیک) در نظر گرفت. میدان‌های تصادفی شرطی نیز از دسته مدل‌های تفکیکی می‌باشند. علاوه بر این، در مقایسه با مدل‌های مارکوف مخفی، میدان‌های تصادفی شرطی، ملزم به داشتن ساختار دنباله خطی نمی‌باشند و می‌توانند به طور دلخواه ساختار بندی شوند.

همان طور که مطرح شد، تفاوت اصلی مدل‌های مولد و تفکیکی در این است که در توزیع شرطی $p(y|x)$ ، مدل $p(x)$ مورد نیاز برای طبقه‌بندی وجود ندارد. مشکل مدل‌سازی $p(x)$ در این است که این مدل اغلب دارای ویژگی‌های وابسته بسیاری می‌باشد که مدل‌سازی آن‌ها دشوار می‌باشد. به عنوان مثال، کاربرد ساده مدل مارکوف مخفی در تشخیص موجودیت‌های اسمی، تنها بر یک ویژگی یعنی شناسایی (هویت) کلمه تکیه می‌کند. اما بسیاری از کلمات، به ویژه نام‌های خاص، در مجموعه آموزشی وجود ندارند، بنابراین ویژگی شناسایی (هویت) کلمه ناکارآمد واقع می‌شود. قصد داریم برای برچسب‌گذاری کلمات مشاهده نشده، از سایر ویژگی‌های کلمه، از قبیل حروف بزرگ آن، کلمات مجاور آن، پیشوندها و پسوندهای آن، عضویت در فهرست‌های از پیش تعیین شده افراد و مکان‌ها و سایر موارد استفاده نماییم. مزیت اصلی مدل‌سازی تفکیکی در این است که این مدل برای در نظر گرفتن ویژگی‌های غنی و دارای اشتراک مناسب هستند؛

اما چنین قابلیت در مدل مارکوف وجود ندارد.

۲-۳- چند مثال از گراف بدون جهت

از آنجا که میدان تصادفی شرطی زیرمجموعه مدال‌های گرافی بدون جهت است. مثال‌هایی از این روش‌ها برای درک بهتر مدل‌ها ارائه خواهیم کرد.

۲-۳-۱- سناریوی دانشجویان متقلب

مثال ساده‌ای از گراف احتمالی بدون جهت برای مدل‌سازی تقلب دانشجویان ارائه شده است. فرض کنید چهار دانشجو به نام‌های کتی، هیدر، مارک و آلیسن باید در دو آزمون شرکت کنند. در آزمون اول هر دانشجو در کلاس مجزا است. چون هیدر و آلیسن مطالعات زیادی داشته‌اند به ۹۰٪ سوالات پاسخ صحیح می‌دهند. از سوی دیگر کتی و مارک مطالعه چندانی نداشته‌اند و ۲۵٪ سوالات را صحیح پاسخ می‌دهند. (فرض بر این است که سطح همه سوالات باهم برابر است.)

در آزمون دوم همه دانشجویان در یک کلاس قرار دارند. آنها به ترتیب زیر نشسته‌اند.

آلیسن - مارک - هیدر - کتی

آنها می‌توانند ورقه کناردستی خود را مشاهده کنند. هر چهار نفر گمان می‌کنند که فرد کناردستی برای آزمون مطالعه کرده است و برای هر سوال احتمالاً گزینه صحیح را انتخاب می‌کند. در این حالت پاسخ هر دانشجو به دانشجوی دیگر وابسته می‌شود؛ زیرا هر دانشجو ورقه کناردستی خود را می‌بیند و گمان می‌کند که او به سوالات پاسخ صحیح می‌دهد. حتی پاسخ کتی و مارک هم به هم وابسته است، چون هر دو ورقه هیدر را می‌بینند و به اصطلاح مسیری در گراف برای اتصال بین آن دو وجود دارد. اگرچه هیچکدام از چهار متغیر مستقل نیستند. اما می‌توان آنها را به صورت شرطی مستقل دید. به طور مثال اگر ما پاسخ هیدر را بدانیم؛ در آن صورت پاسخ مارک و کتی از هم مستقل خواهد بود. در مدل گرافی می‌گوییم دو متغیر به صورت شرطی مستقل هستند اگر حذف متغیر شرطی از گراف باعث شود دو متغیر از هم مستقل شوند. همان‌طور که گفته شد در گراف بالا اگر هیدر حذف شود مسیری بین مارک و کتی وجود نخواهد داشت.

اگر تابع توزیع احتمال دلخواه $p(\mathbf{y})$ بتواند با ضرب فاکتورهایی به فرم $\Psi_a(\mathbf{y}_a)$ نشان داده شود؛ که در آن a شمارنده ایست که به یک عامل اشاره دارد و در بازه‌ای بین ۱ تا A (تعداد عامل‌ها) تغییر می‌کند [۸]. براساس مدل گرافی بدون جهت، می‌توان تابع توزیع p را به صورت زیر نوشت:

$$p(\mathbf{y}) = Z^{-1} \prod_{a=1}^A \Psi_a(\mathbf{y}_a) \quad (9-2)$$

ثابت Z یک مقدار نرمال‌سازی است و باعث می‌شود که حاصل جمع توزیع یک شود و به صورت زیر محاسبه می‌گردد.

$$Z = \sum_{\mathbf{y}} \prod_{a=1}^A \Psi_a(\mathbf{y}_a) \quad (10-2)$$

باید توجه نمود که محاسبه این حاصل جمع برای مقدارهای مختلف \mathbf{y} نمایی است و در حالت کلی امکان‌پذیر نیست. برای حل مساله تقلب از رابطه (۹-۲) استفاده شده است.

$$p(y_1, y_2, y_3, y_4) = \frac{1}{Z} \prod_{i=1}^4 \Psi_i(y_i) \prod_{e=1}^3 \Psi_e(y_e, y_{e+1}) \quad (11-2)$$

در این مساله یک تابع پتانسیل با نام Ψ_i برای هر دانشجو وجود دارد و یک تابع عامل نیز برای دانشجو کنار هم با نام، Ψ_e وجود دارد (آلیسن - مارک، مارک - هیدر، هیدر - کتی).

در مساله تقلب مقدار تابع پتانسیل Ψ_i ، برابر دقت پاسخ‌گویی هر فرد به صورت انفرادی در نظر گرفته شده است. مقدار تابع پتانسیل Ψ_e در شرایطی که دو فرد کنار دست هم یک نظر داشته باشند دو برابر شرایطی که نظر متفاوت داشته باشند فرض شده است. ثابت Z برای آن است که تابع توزیع احتمال دارای جمع یک باشد و قوانین احتمال برای آن صادق باشد. برای حل مساله تمام حالات ممکن لیست شده است و در جدول (۱-۲) آورده شده است. احتمال پاسخ صحیح از جانب مارک ۰,۲۵ است. چون کمیتی به نام ثابت Z وجود دارد. می‌توان به جای ۰,۲۵ نسبت آن را رعایت نمود. از این‌رو در جایی که پاسخ صحیح داده است فاکتور ۱ در نظر گرفته شده و در جایی که پاسخ غلط داده شده است فاکتور ۳ در نظر گرفته شده است. برای محاسبه ثابت Z مقادیر ستون «ضرب عامل‌ها» باهم جمع شده است (عدد ۳۷۹۰) و برای یافتن احتمال رخداد حالت‌ها، مقادیر ستون «ضرب عامل‌ها» بر عدد ثابت Z تقسیم شده است.

حالتی که تمام دانشجویان جواب درست می‌دهند؛ دارای احتمال ۰,۱۷ است، این در شرایطی است که احتمال آنکه این چهار دانشجو در امتحان انفرادی پاسخ درست دهند، ۰,۰۵ است.

همان‌طور که مشاهده می‌شود برای یافتن ثابت Z باید تمام حالات لیست شوند و در حالت عمومی محاسبه

آن دست‌نیافتنی است. در این مساله صرفاً متغیرهای حالت حضور داشتند و مشاهده‌ای وجود نداشت.

جدول (۲-۱) تمام حالت‌های ممکن برای مساله تقلب

احتمال	ضرب عامل‌ها	Ψ_{e3}	Ψ_{e2}	Ψ_{e1}	Ψ_{i4}	Ψ_{i3}	Ψ_{i2}	Ψ_{i1}	کتی	هیدر	مارک	آلیسن
۰,۱۷	۶۴۸	۲	۲	۲	۹	۹	۱	۹	ص	ص	ص	ص
۰,۲۶	۹۷۲	۱	۲	۲	۹	۹	۱	۹	غ	ص	ص	ص
۰,۰۰	۱۸	۱	۱	۲	۱	۱	۱	۹	ص	غ	ص	ص
۰,۰۳	۱۰۸	۲	۱	۲	۱	۱	۱	۹	غ	غ	ص	ص
۰,۱۳	۴۸۶	۲	۱	۱	۹	۹	۳	۹	ص	ص	غ	ص
۰,۱۹	۷۲۹	۱	۱	۱	۹	۹	۳	۹	غ	ص	غ	ص
۰,۰۱	۵۴	۱	۲	۱	۱	۱	۳	۹	ص	غ	غ	ص
۰,۰۹	۳۲۴	۲	۲	۱	۱	۱	۳	۹	غ	غ	غ	ص
۰,۰۱	۳۶	۲	۲	۱	۹	۹	۱	۱	ص	ص	ص	غ
۰,۰۱	۵۴	۱	۲	۱	۹	۹	۱	۱	غ	ص	ص	غ
۰,۰۰	۱	۱	۱	۱	۱	۱	۱	۱	ص	غ	ص	غ
۰,۰۰	۶	۲	۱	۱	۱	۱	۱	۱	غ	غ	ص	غ
۰,۰۳	۱۰۸	۲	۱	۲	۹	۹	۳	۱	ص	ص	غ	غ
۰,۰۴	۱۶۲	۱	۱	۲	۹	۹	۳	۱	غ	ص	غ	غ
۰,۰۰	۱۲	۱	۲	۲	۱	۱	۳	۱	ص	غ	غ	غ
۰,۰۲	۷۲	۲	۲	۲	۱	۱	۳	۱	غ	غ	غ	غ

۲-۳-۲- حرفه‌های دانش‌آموختگان علوم کامپیوتری

مثال حرفه دانش‌آموختگان، یک مدل گرافی بدون جهت زنجیره‌ای است. این مثال یک مدل زنجیره‌ای مارکوف^۱ است که نشان می‌دهد بعد از دانش‌آموختگی در رشته علوم کامپیوتر افراد چه شغلی انتخاب

می‌کنند. در ابتدا هر دانش‌آموخته با یکی از سه شغل زیر شروع می‌کند.

جدول (۲-۲) احتمال رخداد حالت اولیه

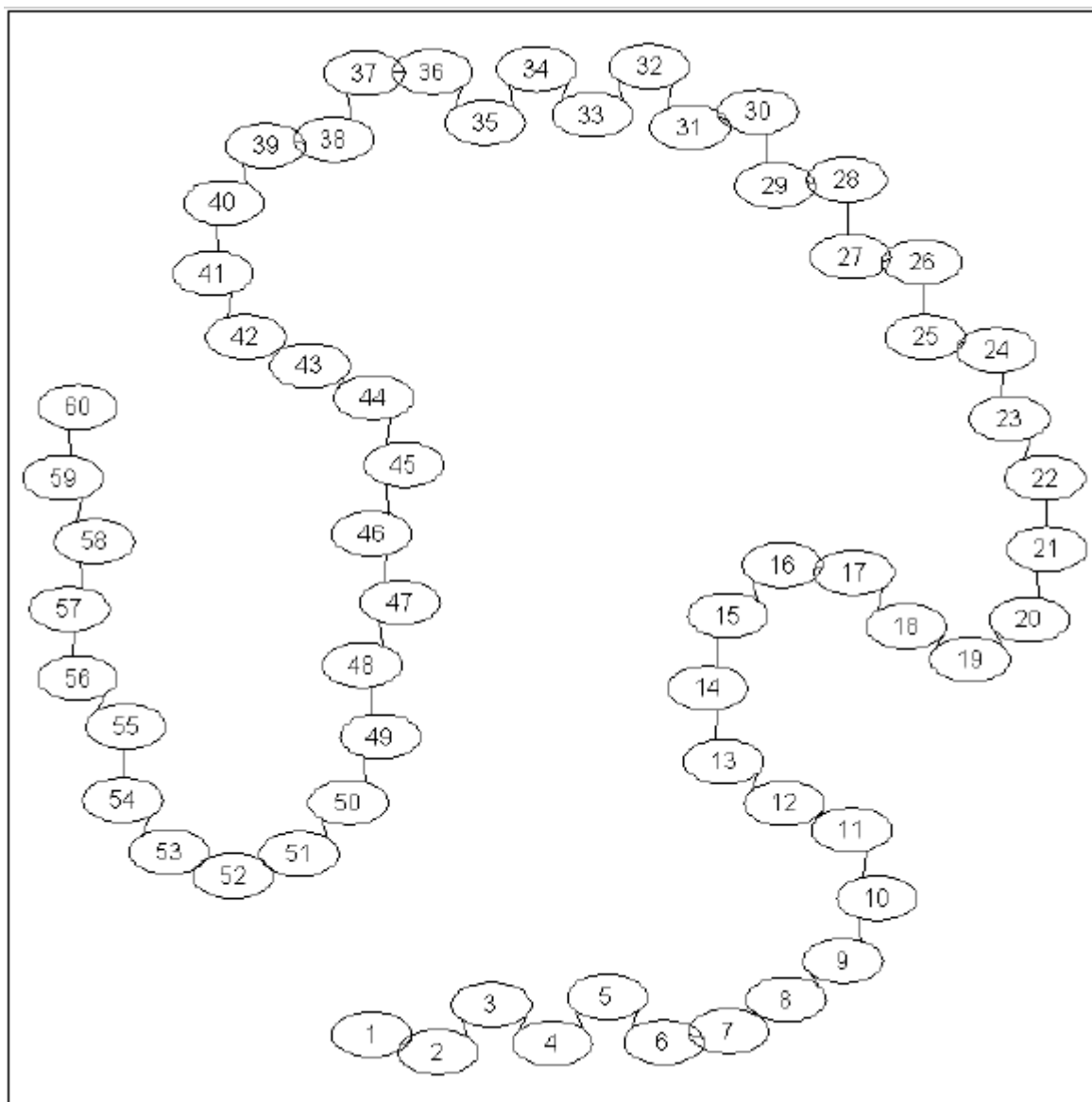
وضعیت	احتمال
صنعت	۰,۶
ادامه تحصیل (در مقطع کارشناسی ارشد یا دکتری)	۰,۳
بازی ویدئویی	۰,۱

پس از سال اول، فرد تصمیم می‌گیرد که در همان شغل ادامه دهد یا آنکه شغل دیگری انتخاب کند. در سال‌های بعد چهار شغل دیگر هم به انتخاب‌های او اضافه می‌شود که عبارتند از: صنعت (با درجه دکتری)، هیات علمی، بازی ویدئو (با درجه دکتری) و خاتمه فعالیت به عنوان متخصص کامپیوتر. ماتریس تبدیل وضعیت در زیر آمده است.

جدول (۳-۲) ماتریس فضای حالت

از	بازی ویدئویی	صنعت	ادامه تحصیل	بازی... (دکتری)	صنعت (دکتری)	هیات علمی	خاتمه
بازی ویدئویی	۰,۰۸	۰,۰۹	۰,۰۱	۰	۰	۰	۰,۰۱
صنعت	۰,۰۳	۰,۹۵	۰,۰۱	۰	۰	۰	۰,۰۱
ادامه تحصیل	۰,۰۶	۰,۰۶	۰,۷۵	۰,۰۵	۰,۰۵	۰,۰۲	۰,۰۱
بازی ویدئویی (دکتری)	۰	۰	۰	۰,۳	۰,۶	۰,۰۹	۰,۰۱
هیات علمی	۰	۰	۰	۰,۰۱	۰,۰۱	۰,۹۷	۰,۰۱
خاتمه	۰	۰	۰	۰	۰	۰	۱

از آنجا که این مساله یک مدل مارکوفی است پس شغل امسال دانش‌آموخته تنها به شغل سال گذشته آن وابسته است. شکل (۲-۴) این مدل را برای ۶۰ سال شغل دانش‌آموخته نشان می‌دهد. این مساله یک ارتباط زنجیره‌ای است که هم با مدل مارکوف و هم با میدان تصادفی شرطی خطی قابل حل است. اعداد مساله به صورت توابع توزیع احتمال هستند؛ به همین دلیل می‌توان Z را برابر یک در نظر گرفت. حل این مساله با الگوریتم برنامه‌نویسی پویای ویتربی و الگوریتم پیش‌رو-پس‌رو در دسترس است که توضیحات آن در گذشته آمد.



شکل (۲-۴) وضعیت شغل دانش‌آموخته در طی ۶۰ سال

مثال‌هایی دیگری در ساختارهای گرافی بدون جهت وجود دارد که در آنها مدل گراف حالت به صورت درخت است و یا در گراف دور وجود دارد که برای می‌توان برای فهم بیشتر به [۸] مراجعه نمود.

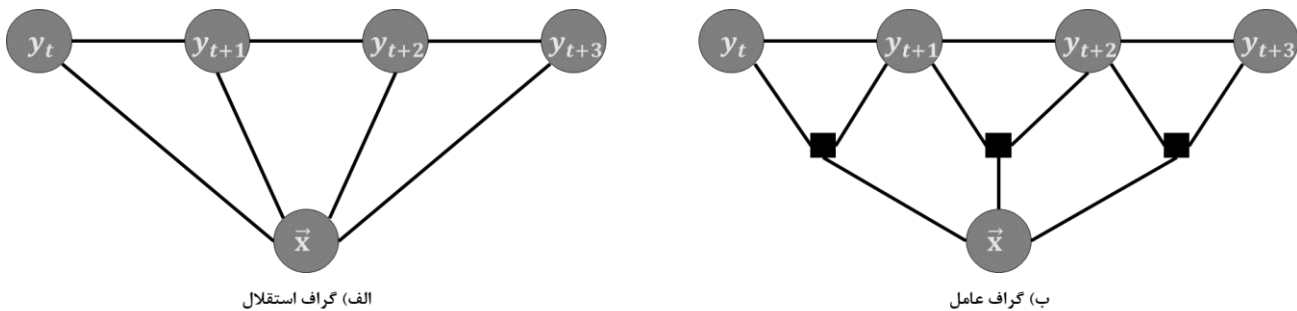
۲-۴- میدان تصادفی شرطی زنجیره‌ای خطی

شکل خاصی از یک میدان تصادفی شرطی که دارای ساختار زنجیره خطی است و متغیرهای خروجی را به صورت یک دنباله مدل می‌کند. شکل (۲-۵) گراف استقلال و گراف عامل معادل با میدان تصادفی شرطی

خطی را نشان می دهد.

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{t=1}^T \psi_t(y_t, y_{t-1}, \mathbf{x}_t) \quad (12-2)$$

$$\Psi_t(y_t, y_{t-1}, \mathbf{x}_t) = \exp \left\{ \sum_{k=1}^K \theta_k f_k(y_t, y_{t-1}, \mathbf{x}_t) \right\} \quad (13-2)$$



شکل (۲-۵) یک میدان تصادفی شرطی زنجیره‌ی خطی [۲]

در میدان‌های تصادفی شرطی زنجیره‌ی خطی، عامل‌های زبانی به صورت رابطه (۲-۱۳) در نظر گرفته می شوند. طول دنباله T است و θ_k ضریب ویژگی k ام را نشان می دهد که باید از روی داده‌های ورودی محاسبه شود.

برای به کارگیری روش میدان تصادفی شرطی، متناسب با مسئله باید تعدادی تابع ویژگی تعریف کرد. نقش توابع ویژگی، تشویق الگوهای مناسب و تنبیه الگوهای نامناسب برچسب گذاری است. در واقع سعی می کنیم برچسب گذاری‌های دلخواه در دنباله را در قالب توابع ویژگی بیان کنیم. از رابطه‌ی بالا مشخص است که توابع ویژگی، تابعی از نمونه \mathbf{x}_t ، برچسب احتمالی آن و برچسب احتمالی نمونه قبل از آن می باشد. معمولاً به جهت سادگی، توابع ویژگی دودویی طراحی می شوند اما در حالت کلی می توانند هر مقدار حقیقی داشته باشند.

به عنوان مثال در مسئله برچسب گذاری واژگان در زبان انگلیسی، توابع زیر با توجه به دستور زبان انگلیسی قابل تعریف می باشند:

$$(14-2)$$

$$f_1(x_i, y_i, y_{i-1}) = \begin{cases} 1 & \text{if } y_{i-1} = \text{ADJECTIVE and } y_i = \text{NOUN} \\ 0 & \text{else} \end{cases}$$

(۱۵-۲)

$$f_2(x_i, y_i, y_{i-1}) = \begin{cases} 1 & \text{if } y_{i-1} = \text{PREPOSITION and } y_i = \text{PREPOSITION} \\ 0 & \text{else} \end{cases}$$

$$f_3(x_i, y_i, y_{i-1}) = \begin{cases} 1 & \text{if } y_i = \text{ADVERB and } x_i \text{ ends with "ly"} \\ 0 & \text{else} \end{cases} \quad (۱۶-۲)$$

تابع f_1 به برچسب‌گذاری‌ای که اسم پس از صفت بیاید امتیاز مثبت می‌دهد. تابع f_2 به برچسب‌گذاری‌ای که در آن دو نمونه پشت سرهم «حرف اضافه» تشخیص داده شوند امتیاز منفی می‌دهد (منفی بودن این امتیاز از طریق ضرایب θ در گام یادگیری لحاظ خواهد شد). به چنین توابعی که تنها به برچسب نمونه‌ها وابسته هستند و نه خود نمونه ویژگی لبه‌ای^۱ (جفتی) گفته می‌شود. واضح است که هر چه تعداد این ویژگی‌ها بیشتر باشد، دسته‌بندی بیشتر وابسته به دانش قبلی به دست آمده از مسئله و مستقل از داده‌های حاضر می‌شود.

تابع f_3 ، یک نمونه تابع ویژگی را نشان می‌دهد که در آن از ویژگی‌های نمونه‌ی نام استفاده شده است. در این تابع برچسب‌گذاری «قید» برای کلماتی که پسوند «ly» دارند، تشویق شده است. به چنین توابعی که تنها به یک نمونه و برچسب آن وابسته هستند، ویژگی گره‌ای^۳ (تکی) گفته می‌شود. این توابع در واقع متناظر با همان نوع نگاه مستقل به هر داده در الگوشناسی آماری می‌باشند. این‌ها، تنها مثال‌های از انواع توابع ویژگی می‌باشند که در طول زمان به صورت الگو در آمده‌اند [۹]. اما به‌طور کلی توابع ویژگی طبق رابطه (۲-۱۳) می‌توانند در یک لحظه به هر دو برچسب نمونه‌های حال و قبلی و همین‌طور ویژگی‌های نمونه حاضر وابسته باشند.

آموزش و استنتاج^۶

برای به‌کارگیری میدان تصادفی شرطی باید دو مسئله آموزش و استنتاج حل شود. در مرحله آموزش پس از انتخاب توابع ویژگی مناسب هدف یافتن پارامترهای θ_i مناسب به ازای داده‌های آموزشی است. این عمل از طریق بیشینه شدن $p(y|x)$ با توجه به داده‌های آموزشی X و Y انجام می‌گیرد. روش‌های مختلفی برای

۱- Edge Feature

۲- Pairwise Feature

۳- Node Feature

۴- Unary Feature

۵- Training

۶- Inference

این منظور ارائه شده است. به طور معمول آموزش یک میدان تصادفی شرطی با استفاده از روش تخمین بیشینه احتمال انجام می‌شود [۱۰][۱۱]. این مساله متناظر با مساله سوم در مدل مارکوف مخفی است که توضیحات آن بیان شد.

$$\theta^* = \arg \max_{\theta} p_{\theta}(Y|X) \quad (1-1)$$

اما این روش نیاز به محاسبه‌ی $Z(x)$ دارد که یک مسئله‌ی NP می‌باشد. یک راه دیگر برای تخمین احتمال، جایگزین کردن شبه احتمال^۲ به جای احتمال می‌باشد. آموزش در روش شبه احتمال تنها به توزیع شرطی متغیرهای منفرد وابسته می‌باشد. از آن جا که ثابت نرمال سازی برای این توزیع فقط به متغیرهای تک بستگی دارد، می‌توان آن را به صورت موثری محاسبه کرد.

بعد از آموزش میدان تصادفی شرطی و دستیابی به وزن‌های θ_i ، می‌توان از این مدل در جهت تشخیص و طبقه‌بندی داده‌های جدید استفاده کرد. یک استنتاج کارآمد برای میدان‌های تصادفی شرطی بسیار حیاتی و مهم می‌باشد، چرا که در هر دو مرحله‌ی آموزش و پیش‌بینی برچسب‌های ورودی‌های جدید، نیاز به استنتاج می‌باشد. در هر دو روش مدل مارکوف مخفی و میدان تصادفی شرطی، پاسخ بهینه مسئله را می‌توان به کمک برنامه‌نویسی پویا و با الگوریتمی به نام ویتربی، در زمان چند جمله‌ای محاسبه کرد. به عبارتی این موضوع متناظر با مساله دوم مدل مارکوف مخفی است.

۱- Maximum Likelihood Estimation

۲- Pseudo-likelihood

فصل سوم

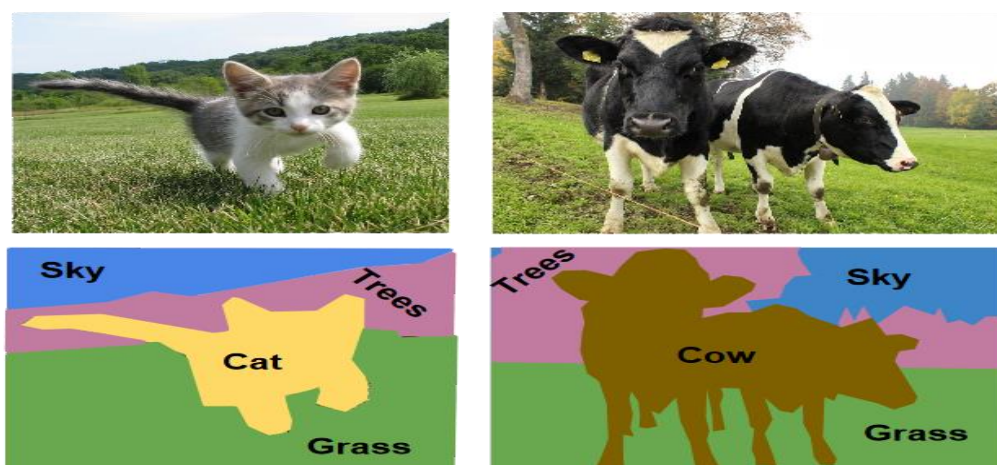
کاربردهای میدان تصادفی شرطی

۳-۱- کاربردهای میدان تصادفی شرطی

میدان‌های تصادفی شرطی یکی از روش‌های مدل‌سازی احتمالی می‌باشد که در سال‌های اخیر توجه زیادی را به خود جلب کرده است و کاربردهای فراوانی را پیدا کرده است. در پردازش زبان طبیعی و زیر شاخه های آن، CRF به‌عنوان روشی بدیع و کارا مطرح شده است. از جمله این کارها می‌توان به برچسب‌گذاری کلمات [۷]، قطعه‌بندی متن [۱۲]، تشخیص موجودیت‌های اسمی [۱۳] و سایر موارد اشاره کرد. علاوه بر این کاربردهای فراوانی از CRF در پیش‌بینی ژن [۱۴]، تشخیص حمله در شبکه [۱۵] و مدیریت سنسورها [۱۶] ارائه شده است. دامنه حضور CRF به همین محدود نمی‌شود و از آنجایی که CRF‌ها در حالت گرافیکی بدون جهت هستند، قابلیت مدل‌سازی خصوصیات تصویری را تا حد زیادی دارا می‌باشند [۱۷]. بنابراین CRF می‌تواند در کاربردهای تصویری مانند برچسب‌گذاری معنایی تصویر، تشخیص شی در تصاویر، تشخیص برجستگی تصویر و سایر موارد نیز مورد استفاده قرار گیرد. ما در این بخش قصد داریم تعدادی از کاربردهای CRF را در تصویر بررسی کنیم.

۳-۲- میدان تصادفی شرطی در قطعه‌بندی معنایی تصویر

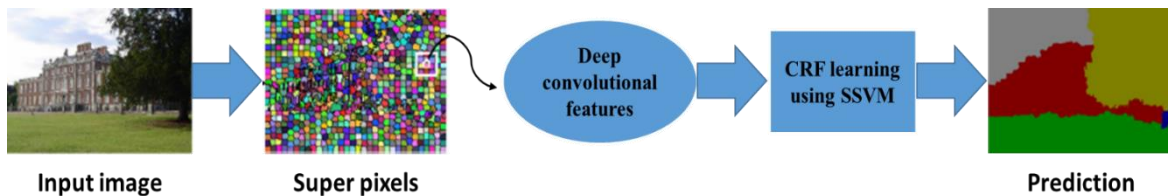
یکی از کاربردهای مهم میدان تصادفی شرطی در قطعه‌بندی معنایی تصاویر می‌باشد. در این مسائل هدف برچسب زدن به هر یک از نواحی تصویر براساس ماهیت و دیگر خواص این نواحی می‌باشد. قطعه‌بندی معنایی تصاویر می‌تواند برای اهداف دیگر، مانند جستجو برای پایگاه داده تصویر (مثلاً «پیدا کردن همه تصاویر با حیوانات در آب») مفید باشد. شکل (۳-۱) نمونه‌ای از قطعه‌بندی معنایی تصاویر را نشان می‌دهد. در ادامه به بررسی تعدادی مقاله در این زمینه می‌پردازیم.



شکل (۳-۱) مثالی از قطعه‌بندی معنایی تصاویر [۱۸]

۳-۲-۱- یادگیری CRF با ویژگی‌های CNN برای قطعه‌بندی معنایی تصویر

Liu و همکاران (۲۰۱۵) [۱۹]، CRF مبتنی بر SSVM را براساس ویژگی‌های یاد گرفته شده از یک شبکه عصبی کانولوشنی عمیق که از قبل آموزش دیده را برای قطعه‌بندی تصویر پیشنهاد می‌کنند. در راستای یادگیری CRF، یافتن یک نمایش خوب ویژگی خیلی مهم است و می‌تواند اثر بنیادی را روی دقت قطعه‌بندی داشته باشد. Liu و همکاران معتقد بودند اغلب مطالعات گذشته بر روش‌های بدون ناظر استخراج ویژگی مانند روش‌های هیستوگرام‌های رنگی، توصیفگرهای HOG یا SIFT جهت ساختن ویژگی‌ها، تکیه می‌کنند. اخیراً یادگیری ویژگی و مخصوصاً روش‌های یادگیری عمیق محبوبیت زیادی را در یادگیری ماشین و زمینه‌های مرتبط به دست آورده‌اند. بنابراین آن‌ها استفاده از یک شبکه‌ی عصبی کانولوشنی (CNN) بزرگ که از قبل آموزش داده شده را برای ایجاد ویژگی‌های عمیق برای یادگیری CRF پیشنهاد می‌کنند. CNN عمیق در پایگاه داده‌ی ImageNet آموزش داده می‌شود. علاوه بر این Liu و همکاران یک سری پتانسیل‌های جفتی جدید که نشان‌دهنده ارتباط مکانی نواحی مختلف می‌باشد را ایجاد می‌کنند. شکل (۲-۳) خط مشی روش قطعه‌بندی را نشان می‌دهد که در آن ابتدا تصویر به سوپرپیکسل‌ها تقسیم و سپس ویژگی‌های کانولوشنی عمیق با استفاده از یک شبکه عمیق کانولوشنی آموزش داده می‌شود. در نهایت ویژگی‌های آموزش داده شده را با استفاده از یادگیری CRF قطعه‌بندی می‌کنیم.



شکل (۲-۳) مراحل روش قطعه‌بندی پیشنهادی [۱۹]

Liu و همکاران با توجه به نمونه تصویر x و برجسب مربوطه به آن y ، CRF را به صورت زیر مدل می‌کنند:

$$p(y|x; w) = \frac{1}{Z} \exp(-E(y, x; w)) \quad (1-3)$$

که در آن w پارامترها و Z معیار نرمال شده می‌باشد. انرژی E تصویر x با برجسب قطعه‌بندی y بر روی گره‌ها (سوپرپیکسل‌ها) N و لبه‌های S به شکل زیر محاسبه می‌شود:

$$E(y, x; w) = \sum_{p \in N} \varphi^{(1)}(y^p, x; w) + \sum_{(p, q) \in S} \varphi^{(2)}(y^p, y^q, x; w) \quad (2-3)$$

در اینجا $x \in X$ و $y \in Y$ ، $\varphi^{(1)}$ و $\varphi^{(2)}$ پتانسیل‌های انحصاری و جفتی هستند که هر دو به مشاهدات و همچنین پارامتر w بستگی دارند.

منظور از مفهوم پتانسیل جفتی این است که یک سری جفت مناطق هستند که در تصاویر مختلف کنار هم قرار می‌گیرند و بعضی جفت مناطق منحصر به فرد هستند که کمتر احتمال دارد که با هم ظاهر شوند. برای مثال احتمال اینکه گاو با چمن‌زار ظاهر شود بیشتر از این است که گاو با نمایشگر ظاهر شود یا احتمال ظاهر شدن آسمان در بالای تصویر خیلی زیاد است تا اینکه در پایین تصویر ظاهر شود. اما پتانسیل های جفتی که Liu و همکاران در نظر گرفته‌اند چهار رابطه فضایی را بررسی می‌کند. چهار رابطه فضایی که در نظر گرفته می‌شود عبارتند از: p بالای q است، p زیر q است، p در سمت چپ q است و p سمت راست q است. ترسیم این ویژگی برای پتانسیل جفتی به صورت رابطه (۳-۳) نوشته می‌شود:

$$\sum_{(p,q) \in S} \varphi^{(2)}(y^p, y^q, x) = \sum_{(p,q) \in S_1} \varphi_1^{(2)}(y^p, y^q, x) + \sum_{(p,q) \in S_2} \varphi_2^{(2)}(y^p, y^q, x) + \sum_{(p,q) \in S_3} \varphi_3^{(2)}(y^p, y^q, x) + \sum_{(p,q) \in S_4} \varphi_4^{(2)}(y^p, y^q, x) \quad (3-3)$$

زمانی که p و q در رابطه‌های فضایی هستند، S_1, S_2, S_3 و S_4 مجموعه‌های کناره، به ترتیب «بالا»، «پایین»، «چپ» و «راست» و $S = S_1 \cup S_2 \cup S_3 \cup S_4$ و $S_i \cap S_j = \emptyset$ برای $i \neq j$ و $i, j = 1, 2, 3, 4$ می‌باشند. برای ایجاد پتانسیل‌های جفتی وقوع همزمان، در نظر می‌گیریم که داده‌های آموزشی به اندازه کافی بزرگ هستند. پتانسیل‌های جفتی می‌توانند به صورت رابطه (۴-۳) نوشته شوند.

$$\varphi_i^{(2)}(y^p, y^q, x) = L_{pq} \cdot I(y^p \neq y^q) \cdot g_i(y^p, y^q) \cdot i = 1, 2, 3, 4 \quad (4-3)$$

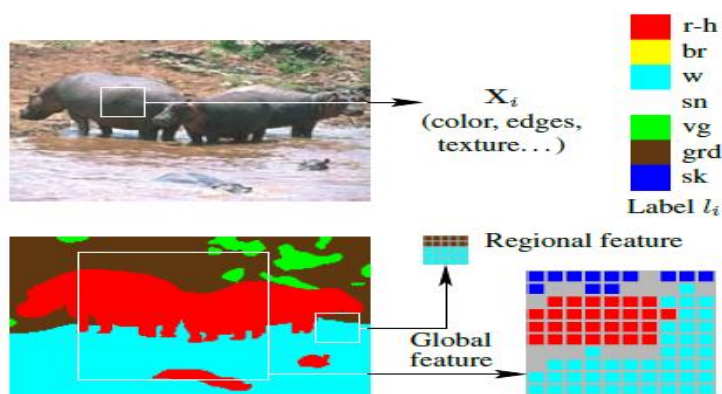
که در آن $g_i(y^p, y^q) = \frac{1}{f_{co-occur}^i(y^p, y^q)}$ و $f_{co-occur}^i(y^p, y^q) = \frac{N_{pq}^i}{N_{pq}}$ می‌باشد. در اینجا N_{pq} تعداد تصاویر آموزشی که y^p و y^q با هم موجود باشند و N_{pq}^i ($i = 1, 2, 3, 4$) تعداد تصاویر آموزشی هستند که y^p و y^q به ترتیب در ابرپیکسل مجاور مرتبط با چهار رابطه‌ی فضایی ظاهر می‌شوند. اگر $N_{pq}^i = 0$ باشد، بدین معنی است که y^p و y^q هرگز در i امین رابطه فضایی ظاهر نمی‌شود، سپس $g_i(y^p, y^q) = \inf$ که استنتاج از دستیابی به چنین زوج برچسب‌گذاری ممانعت به عمل می‌آورد. بصورت خلاصه، Liu و همکاران نشان دادند که:

- ویژگی‌های متفاوتی که از تصاویر توسط CNNها از داده‌های برچسب‌گذاری شده در مجموعه داده ImageNet به دست می‌آیند را می‌توان به صورت موفقیت آمیزی برای هدف قطعه‌بندی استفاده کرد.
- CRF مبتنی بر SSVM با ویژگی CNN نتایج شگفت‌انگیزی را ایجاد می‌کند.
- پتانسیل‌های جفتی وقوع همزمان مرتبط با چهار فضا را درون استنتاج سازماندهی کردند و باعث ارتقای بیشتر عملکرد شده‌اند.

۳-۲-۲- میدان‌های تصادفی شرطی چند مقیاسی برای برچسب‌گذاری تصویر

He و همکاران (۲۰۰۴) [۲۰] یک رویکرد شامل ویژگی‌های متنی برای برچسب‌گذاری تصویر، که هر پیکسل به یک مجموعه محدود از برچسب‌ها اختصاص دارد را پیشنهاد می‌کنند. این ویژگی‌ها در یک چارچوب احتمالی که خروجی‌های چندین اجزا (مولفه) را ترکیب می‌کنند، گنجانده شده‌اند. اجزا در اطلاعاتی که آن‌ها کدگذاری می‌کنند، متفاوت هستند. برخی روی نگاشت برچسب-تصویر، تمرکز می‌کنند، در حالی که برخی فقط روی الگوهای درون فیلد برچسب تمرکز می‌کنند. اجزا نیز در مقیاس، متفاوت هستند، به‌عنوان مثال برخی روی الگوها با وضوح خوب تمرکز می‌کنند در حالی که برخی دیگر روی ساختار درشت‌تر و سراسری تمرکز می‌کنند.

CRF‌های استاندارد از دو شکل توابع ویژگی استفاده می‌کنند که در تصویر دو بعدی به شرح زیر تعریف می‌شوند: توابع ویژگی حالت $f(l_i, X, i)$ برچسب در موقعیت i و تصویر مشاهده شده؛ و توابع ویژگی انتقال $f(l_i, l_j, X, i)$ تصویر و برچسب‌ها در موقعیت i و موقعیت مجاور j در تصویر. He و همکاران این را برای ویژگی‌های برچسب که الگوهای خاصی را در یک زیر مجموعه از متغیرهای برچسب کدگذاری می‌کنند، توسعه می‌دهند. ویژگی‌های برچسب یک شکل از توابع پتانسیل هستند، کدگذاری یک محدودیت خاص بین تصویر و برچسب‌ها در یک منطقه از تصویر را توابع پتانسیل می‌گویند. مثال‌ها در شکل (۳-۳) نشان داده شده‌اند. در اینجا، ویژگی برچسب کوچکتر (منطقه‌ای)، الگوی پیکسل‌های زمین بالای پیکسل‌های آب را کدگذاری می‌کند، در حالی که ویژگی برچسب بزرگتر (سراسری)، پیکسل‌های آسمان در بالای تصویر، پیکسل‌های اسب آبی/کرگدن در وسط و پیکسل‌های آب در نزدیکی پایین را کدگذاری می‌کند. ویژگی‌های سراسری می‌توانند در یک وضوح دشت‌تری عمل کنند. مدل He و همکاران این ویژگی‌های برچسب را براساس مجموعه‌ای از تصاویر برچسب شده یاد می‌گیرد.



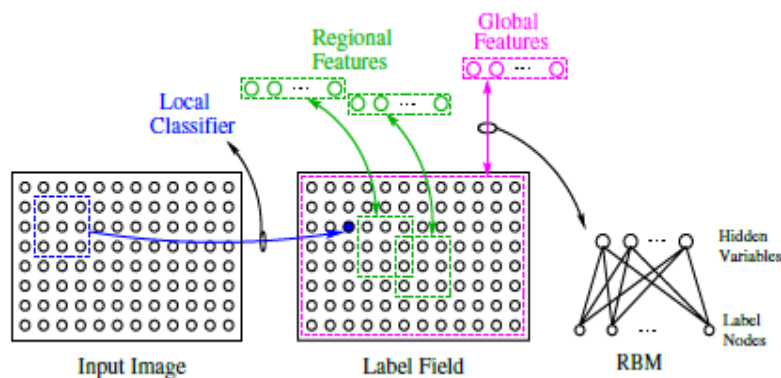
شکل (۳-۳) بالا: یک پیچ تصویر در موقعیت i توسط خروجی چند فیلتر نشان داده شده است. هدف مرتبط کردن پیچ با یکی از مجموعه‌های از پیش تعریف شده برچسب‌ها می‌باشد. پایین: مثال ویژگی‌ها برچسب منطقه‌ای و سراسری [۲۰].

با هر ویژگی برچسب یک متغیر مخفی دودویی همراه می‌شود، که به‌عنوان یک سوئیچ برای این ویژگی عمل می‌کند. این ویژگی یک الگوی برچسب خاص را از طریق یک جدول احتمالی شرطی (CPT) پارامتری شده به موقعیت‌های برچسب در یک منطقه کدگذاری می‌کند. این CPT یک توزیع احتمال چند جمله‌ای را بر روی مقادیر برچسب هر موقعیت مشخص می‌کند. متغیرهای پنهان با توجه به متغیرهای برچسب مربوطه، به صورت شرطی مستقل فرض می‌شوند و برعکس (شکل (۳-۴)). این ساختار به شکل دستگاه بولتزمان محدود (RBM) می‌باشد که در آن استنتاج و یادگیری بسیار ساده شده است.

میدان تصادفی شرطی چند مقیاسی He و همکاران (mCRF) یک توزیع شرطی را بر روی فیلد برچسب $L = \{l_i\}_{i \in S}$ با توجه به تصویر X توسط ترکیب ضربداری مولفه توزیع‌های شرطی که ساختار آماری را در مقیاس‌های مختلف فضایی S ضبط می‌کنند را تعریف می‌کند:

$$P(L|X) = \frac{1}{Z} \prod_s P_s(L|X) \quad (۳-۵)$$

که در آن $Z = \sum_L \prod_s P_s(L|X)$ یک عامل نرمال شده (جمع شده در تمام برچسب‌ها) می‌باشد. بنابراین mCRF یک شکل شرطی از حاصل ضرب مدل‌های خبره می‌باشد. در این مقاله، چارچوب mCRF را با سه جز جداگانه در سه مقیاس مختلف S یک طبقه‌بندی محلی، ویژگی‌های منطقه‌ای و ویژگی‌های سراسری به کار می‌گیرند، همان‌طور که در شکل (۳-۴) نشان داده شده است.



شکل (۳-۴) نمایش مدل گرافیکی. طبقه‌بندی محلی نواحی تصویر را برای برچسب متغیرها طراحی می‌کند، در حالی که متغیرهای مخفی مربوط به ویژگی‌های منطقه‌ای و سراسری یک مدل بدون جهت با برچسب متغیرها را شکل می‌دهند. توجه داشته باشید که ویژگی‌ها و برچسب‌ها به طور کامل متصل هستند و با لایه داخلی متصل نیستند (ماشین بولتزمان محدود) [۲۰].

طبقه‌بندی کننده محلی: یک روش قدرتمند برای طبقه‌بندی یک پیکسل تصویر با استفاده از اطلاعات در سطح محلی تنها استفاده از یک طبقه‌بندی کننده آماری، مانند یک شبکه عصبی می‌باشد. مستقل از هر

موقعیت i ، طبقه‌بندی کننده محلی یک توزیع را روی متغیر برچسب l_i با توجه به خروجی‌های فیلتر x_i در داخل یک تکه تصویر متمرکز شده روی پیکسل i تولید می‌کند.

$$P(L|X, \lambda) = \prod_i P_C(l_i|x_i, \lambda) \quad (6-3)$$

که در آن λ پارامترهای طبقه‌بندی کننده هستند. He و همکاران از یک پرسپکتیو چند لایه به عنوان طبقه‌بند استفاده می‌کنند. توجه داشته باشید که عملکرد طبقه‌بند توسط همپوشانی کلاس و نویز تصویر محدود می‌شود.

ویژگی‌های برچسب منطقه‌ای: این جز دوم برای نشان دادن روابط هندسی محلی بین اشیاء، مانند لبه‌ها، گوشه‌ها یا T-اتصال طراحی شده است. فرض کنید r شاخص مناطق، a شاخص ویژگی‌های منطقه‌ای متفاوت در درون هر منطقه و $\{1, \dots, J\}$ شاخص گره‌های برچسب (موقعیت‌ها) درون منطقه r را نشان می‌دهد. پارامتر $\omega_{a,j}$ ، متغیر منطقه‌ای مخفی $f_{r,a}$ و گره برچسب $l_{r,j}$ را متصل می‌کند. بنابراین $\omega_{a,j}$ می‌تواند به عنوان یک بردار با عناصر $|\mathcal{L}|$ نمایش داده شود. بنابراین، مدل احتمالاتی توصیف ویژگی‌های منطقه‌ای دارای توزیع توام (7-3) می‌باشد.

$$P_R(L, f) \propto \exp \left\{ \sum_{r,a} f_{r,a} W_a^T l_r \right\} \quad (7-3)$$

ویژگی‌های برچسب سراسری: هر ویژگی سراسری با وضوح بالا، حوزه برچسب کل تصویر را در نظر می‌گیرد (شکل (8-3)). این ویژگی‌های سراسری نیز به عنوان یک RBM با اتصال‌های بدون جهت بین متغیرهای سراسری مخفی و متغیرهای برچسب پیکربندی شده‌اند. فرض کنید b الگوهای برچسب سراسری را در پارامترهای $\{u_b\}$ و $g = \{g_b\}$ به عنوان متغیرهای سراسری مخفی دودویی کدگذاری می‌کند. به منظور تشویق این متغیرها برای نشان دادن جنبه‌های درشت فیلد برچسب، He و همکاران فیلد برچسب را به پچ‌های غیر همپوشانی $\{1, \dots, M\}$ تقسیم می‌کنند و هر متغیر سراسری مخفی g_b ، اتصال آن‌ها با گره‌های برچسب درون پچ p_m را به یک بردار پارامتر منفرد u_{b,p_m} اختصاص می‌دهند. همانند مولفه منطقه‌ای، مدل ویژگی برچسب سراسری دارای یک توزیع توام می‌باشد.

$$P_g(L, g) \propto \exp \left\{ \sum_b g_b u_b^T L \right\} \quad (8-3)$$

توزیع احتمال مربوط به مولفه‌ها روی فیلد برچسب در یک فرم بسته ساده به صورت ضرب مولفه‌ها، ترکیب می‌شود (9-3).

$$P(L|X; \theta) = \frac{1}{Z} \prod_i P_C^Y(l_i|x_i, \lambda) \times \prod_{r,a} [1 + \exp(w_a^T l_r)] \times \prod_b [1 + \exp(u_b^T L)] \quad (9-3)$$

که در آن $\theta = \{\lambda, \{w_a\}, \{u_b\}, \gamma\}$ مجموعه‌ای از پارامترهای موجود در مدل می‌باشد.

۳-۳- جمع‌بندی کاربرد میدان تصادفی در حوزه قطعه‌بندی

شبکه‌های عصبی کانولوشن کاربرد موفقی در پردازش تصویر یافته‌اند. از جمله موفقیت‌های این روش استفاده از آن به عنوان ویژگی است. اغلب روش‌های شناسایی و قطعه‌بندی اخیر نیز از آن برای استخراج ویژگی استفاده نموده‌اند. مقاله [۲۱] جدولی ارایه داده است تا نتایج روش‌هایی که شبکه‌های عصبی کانولوشن و میدان تصادفی شرطی را به صورت توأم استفاده می‌کنند را لیست نماید. در این جدول نتایج بر روی مجموعه داده VOC2012 ارایه شده است و نوع شبکه عمیق نشان داده شده و از معیار IoU برای ارزیابی استفاده شده است. در این مقاله روش‌های عمیقی که از میدان تصادفی شرطی برای قطعه‌بندی استفاده می‌کنند به دو دسته بزرگ تقسیم شده‌اند. در دسته اول میدان تصادفی به عنوان یک پس‌پردازش برای شبکه عصبی کانولوشنی در نظر گرفته شده است. دسته دیگر با به کارگیری استنتاج میدان میانگین^۱ آموزش پارامترهای میدان و شبکه را به صورت هم‌زمان با استفاده از یک شبکه پس‌انتشار خطا انجام دادند. میدان میانگین یک روش تکرار شونده است که به طور جدی در بهینه‌سازی کاهش گرادینان تصادفی کاربرد^۲ دارد. از این طریق می‌تواند کارکرد میدان تصادفی شرطی را به عنوان یک شبکه عصبی بازگشتی^۳ در نظر گرفت.

۱- mean-field inference

۲- stochastic gradient descent

۳- recurrent neural network

جدول (۱-۳) بررسی شبکه‌هایی که CRF و CNN را باهم در قطعه‌بندی استفاده می‌کنند.

روش ارایه شده	IoU	نوع شبکه عمیق
O2P[22] روش‌های مرسوم (غیر عمیق)	۴۷,۸	-
SDS[23] FCN[24] Zoom-out[25] روش‌های مبتنی بر CNN بدون CRF	۵۱,۶ ۶۷,۲ ۶۹,۶	AlexNet VGG VGG
Deeplab[26] EdgeNet[27] BoxSup[28] Dilated Conv[29] Centrale Boundries[30] Deeplab Attention[31] LRR[32] Deeplab v2[33] روش‌های مبتنی بر CNN که CRF را در قسمت پس پردازش دارند	۷۱,۶ ۷۳,۶ ۷۵,۲ ۷۵,۳ ۷۵,۷ ۷۶,۳ ۷۹,۳ ۷۹,۷	VGG VGG VGG VGG VGG VGG ResNet ResNet
CRF as RNNs [34] Deep Gaussian CRF [35] Deep parsing network (DPN) [36] Context [37] High-order CRF [38] Deep Gaussian CRF[35] روش‌های مبنی بر CNN که CRF را در یک فرایند انتها به انتها استفاده می‌کنند	۷۴,۷ ۷۵,۵ ۷۷,۵ ۷۷,۸ ۷۷,۹ ۸۰,۲	VGG VGG VGG VGG VGG ResNet

مراجع

- [1] C. Sutton and A. McCallum , "An Introduction to Conditional Random Fields," *Foundations and Trends in Machine Learning*, pp. 267-373, 2012.
- [2] R. Klinger and K. Tomanek, "Classical Probabilistic Models and Conditional Random Fields," Department of Computer Science, Dortmund University of Technology, 2007.
- [3] J. M. Bernardo, M. J. Bayarri and J. O. Berger, "Generative or discriminative? Getting the best of both worlds," *In Bayesian Statistics, Oxford University Press*, pp. 3-23, 2007.
- [4] I. Ulusoy and C. Bishop, "Generative versus discriminative methods for object recognition," in *Computer Vision and Pattern Recognition*, San Diego, 2005.
- [5] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, The MIT Press, 2009.
- [6] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257 – 286, 1989.
- [7] J. Lafferty, A. McCallum and F. C.N. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, 2001.
- [8] <https://www.cs.ubc.ca/~schmidtm/Software/UGM.html>
- [9] R. C. Sam, H. T. Le, T. T. Nguyen and T. M. Trinh, "Relation Extraction in Vietnamese Text Using Conditional Random Fields," *Asia Information Retrieval Symposium*, pp. 330-339, 2010.
- [10] R. Aster, B. Borchers and C. Thurber, *Parameter Estimation and Inverse Problems*, Academic Press, 2013.
- [11] S. N. Vishwanathan and N. Schraudolph, "Accelerated training of conditional random fields with stochastic gradient methods," in *Proceedings of the 23 International Conference*, 2006.
- [12] K. Tomanek, J. Wermter and U. Hahn, "A Reappraisal of Sentence and Token Splitting for Life Sciences Documents," in *Proceedings of the 12th World Congress on Medical Informatics (MEDINFO)*, 2007.
- [13] B. Settles, "Biomedical named entity recognition using conditional random fields and rich feature sets," *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications*, pp. 104-107, 2004.
- [14] D. DeCaprio, J. P. Vinson, M. D. Pearson, P. Montgomery, M. Doherty and J. E. Galagan, "Conrad: Gene prediction using conditional random fields," *Genome Research*, p. 1389–1398,

2007.

[15] K. K. Gupta, B. Nath and K. Ramamohanarao, "Conditional Random Fields for Intrusion Detection," in *Advanced Information Networking and Applications Workshops*, Niagara Falls, 2007.

[16] X. Zhang, D. Aberdeen and S. V. N. Vishwanathan, "Conditional random fields for multi-agent reinforcement learning," in *Proceedings of the 24th International Conference on Machine Learning (ICML)*, Corvallis, 2007.

[17] S. Kumar and M. Hebert, "Discriminative fields for modeling spatial dependencies in natural images," *Proceedings of the 16th International Conference on Neural Information Processing Systems*, pp. 1531-1538, 2003.

[18] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015.

[19] F. Liu, G. Lin and C. Shen, "CRF Learning with CNN Features for Image Segmentation," *Pattern Recognition*, pp. 2983-2992, 2015.

[20] X. He, R. S. Zemel and M. Carreira-Perpinan, "Multiscale conditional random fields for image labeling," in *Computer Vision and Pattern Recognition*, Washington, 2004.

[21] A. Arnab, et al., "Conditional random fields meet deep neural networks for semantic segmentation," *IEEE Signal Processing Magazine, Trans. Pattern Anal. Machine Intell.*, vol. 35, no. 1, pp. 37-52, 2018.

[22] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu, "Free-form region description with second-order pooling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 37, no. 6, pp. 1177–1189, 2015.

[23] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Simultaneous detection and segmentation," in *Proc. IEEE European Conf. Computer Vision*, 2014, pp. 297–312.

[24] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2015, pp. 3431–3440.

[25] M. Mostajabi, P. Yadollahpour, and G. Shakhnarovich, "Feedforward semantic segmentation with zoom-out features," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2015, pp. 3376–3385.

[26] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," in *Proc. Int. Conf. Learning Representations*, 2015.

- [27] L.-C. Chen, J. T. Barron, G. Papandreou, K. Murphy, and A. L. Yuille, "Semantic image segmentation with task-specific edge detection using CNNs and a discriminatively trained domain transform," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2016, pp. 4545–4554.
- [28] J. Dai, K. He, and J. Sun, "Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proc. IEEE Int. Conf. Computer Vision*, 2015, pp. 1635–1643.
- [29] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learning Representations*, 2015.
- [30] I. Kokkinos, "Pushing the boundaries of boundary detection using deep learning," in *Proc. Int. Conf. Learning Representations*, 2016.
- [31] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: Scaleaware semantic image segmentation," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2016, pp. 3640–3649.
- [32] G. Ghiasi and C. C. Fowlkes, "Laplacian pyramid reconstruction and refinement for semantic segmentation," in *Proc. IEEE European Conf. Computer Vision*, 2016, pp. 519–534.
- [33] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," arXiv Preprint, arXiv:1606.00915, 2016.
- [34] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. Torr, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Computer Vision*, 2015, pp. 1529–1537.
- [35] S. Chandra and I. Kokkinos, "Fast, exact and multi-scale inference for semantic image segmentation with deep Gaussian CRFs," in *Proc. IEEE European Conf. Computer Vision*, 2016, pp. 402–418.
- [36] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, "Deep learning Markov random field for semantic segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, to be published.
- [37] G. Lin, C. Shen, A. van den Hengel, and I. D. Reid, "Exploring context with deep structured models for semantic segmentation," arXiv Preprint, arXiv:1603.03183, 2016.
- [38] A. Arnab, S. Jayasumana, S. Zheng, and P. Torr, "Higher order conditional random fields in deep neural networks," in *Proc. IEEE European Conf. Computer Vision*, 2016, pp. 524–540.
- [39] S. Ramanathan, H. Katti, N. Sebe, M. Kankanhalli and T.-S. Chua, "An Eye Fixation Database for Saliency Detection in Images," *11th European conference on Computer vision*, pp. 30-43, 2010.
- [40] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254 - 1259, 1998.

- [41] N. D. Bruce and J. K. Tsotsos, "Saliency Based on Information Maximization," in *Proceedings of the 18th International Conference on Neural Information Processing Systems*, Canada, 2005.
- [42] A. Oliva, A. Torralba, M. Castelhana and J. Henderson, "top-down control of visual attention in object detection," in *International Conference on Image Processing (Cat. No.03CH37429)*, Barcelona, 2003.
- [43] B. Goodrich and I. Arel, "Reinforcement learning based visual attention with application to face detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Providence, 2012.
- [44] K.-c. Yang, C. C. Guest and P. K. Das, "Human visual attention map for compressed video," in *Eighth IEEE International Symposium on Multimedia (ISM'06)*, San Diego, 2006.
- [45] Z. Hua, Y. Li and J. Li, "Image segmentation algorithm based on improved visual attention model and region growing," in *6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, Chengdu, 2010.
- [46] V. Setlur, S. Takagi, R. Raskar, M. Gleicher and B. Gooch, "Automatic image retargeting," in *4th international conference on Mobile and ubiquitous multimedia*, New Zealand, 2005.
- [47] A. Bamidele, F. W. M. Stentiford and J. Morphet, "An attention-based approach to content-based image retrieval," *BT Technology Journal*, pp. 151-160, 2004.
- [48] L. Dupont, K. Ooms, M. Antrop and V. V. Eetvelde, "Comparing saliency maps and eye-tracking focus maps: The potential," *Landscape and Urban Planning*, pp. 17-26, 2016.
- [49] L. Dupont, K. Ooms, M. Antrop and V. V. Eetvelde, "Comparing saliency maps and eye-tracking focus maps: The potential," *Landscape and Urban Planning*, pp. 17-26, 2016.
- [50] T. Foulsham, "Saliency and eye movements in the perception of natural," *PhD thesis, University of Nottingham*, 2008.
- [51] J. Sun, H. Lu and X. Liu, "Saliency Region Detection Based on Markov Absorption Probabilities," *IEEE Transactions on Image Processing*, pp. 1639 - 1649, 2015.
- [52] E. Rahtu, J. Kannala, M. Salo and J. Heikkilä, "Segmenting salient objects from images and videos," *ECCV 2010: Computer Vision – ECCV*, pp. 366-379, 2010.

- [53] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569-582, 2015.
- [54] L. Duan, C. Wu, J. Miao, L. Qing and Y. Fu, "Visual saliency detection by spatially weighted dissimilarity," in *CVPR*, Colorado Springs, 2011.
- [55] R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, "Frequency-tuned salient region detection," in *Computer Vision and Pattern Recognition*, Miami, 2009.
- [56] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Computer Vision and Pattern Recognition*, Minneapolis, 2007.
- [57] C. Guo, Q. Ma and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," in *Computer Vision and Pattern Recognition*, Anchorage, 2008.
- [58] B. Jiang, L. Zhang, H. Lu, C. Yang and M.-H. Yang, "Saliency detection via absorbing markov chain," in *Computer Vision*, Sydney, 2013.
- [59] Q. Yan, L. Xu, J. Shi and J. Jia, "Hierarchical saliency detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, Portland, 2013.
- [60] F. Perazzi, P. Krähenbühl, Y. Pritch and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Conference on Computer Vision and Pattern Recognition*, Providence, 2012.
- [61] Y. Wei, F. Wen, W. Zhu and J. Sun, "Geodesic saliency using background priors," in *Computer Vision*, 2012.
- [62] C. Yang, L. Zhang, H. Lu, X. Ruan and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Computer Vision and Pattern Recognition*, Portland, 2013.
- [63] C. Li, Y. Yuan, W. Cai, Y. Xia and D. D. Feng, "Robust saliency detection via regularized random walks ranking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015.
- [64] Y. Qin, H. Lu, Y. Xu and H. Wang, "Saliency detection via cellular automata," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015.
- [65] A. D. Andrushia and R. Thangarajan, "Center bias enhanced visual saliency detection method," in *Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*, Chennai, 2017.
- [66] Z.-J. Wang, L. Ma, X. Lin and X. Wu, "MSGC: A New Bottom-Up Model for Salient Object Detection," in *IEEE International Conference on Multimedia and Expo (ICME)*, San Diego, 2018.

- [67] J. Yang and M. Yang, "Top-Down Visual Saliency via Joint CRF and Dictionary Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 576-588, 2016.
- [68] S. Lu, V. Mahadevan and N. Vasconcelos, "Learning Optimal Seeds for Diffusion-based Salient Object Detection," in *Computer Vision and Pattern Recognition*, Columbus, 2014.
- [69] L. Mai, Y. Niu and F. Liu, "Saliency aggregation: A data-driven approach," in *Computer Vision and Pattern Recognition*, Portland, 2013.
- [70] N. Tong, H. Lu, X. Ruan and M.-H. Yang, "Salient object detection via bootstrap learning," in *Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015.
- [71] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, p. 273–297, 1995.
- [72] H. Jiang , J. Wang, Z. Yuan, Y. Wu, N. Zheng and S. Li, "Salient Object Detection: A Discriminative Regional Feature Integration Approach," in *Proceedings of the 2015 Conference on Computer Vision and Pattern Recognition*, 2014.
- [73] J. Kim, D. Han, Y.-W. Tai and J. Kim, "Salient region detection via high-dimensional color transform," in *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 2014.
- [74] L. Wang, H. Lu, X. Ruan and M.-H. Yang, "Deep networks for saliency detection via local estimation and global search," in *Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015.
- [75] R. Zhao, W. Ouyang, H. Li and X. Wang, "Saliency detection by multi-context deep learning," in *Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015.
- [76] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015.
- [77] T. Chen, L. Lin, L. Liu, X. Luo and X. Li, "DISC: Deep Image Saliency Computing via Progressive Representation Learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2015.
- [78] Y. Yuan, C. Li, J. Kim, W. Cai and D. D. Feng, "Dense and Sparse Labeling with Multi-Dimensional Features for Saliency Detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 5, pp. 1130-1143, 2016.
- [79] Y. Yuan, C. Li, J. Kim, W. Cai and D. D. Feng, "Reversion correction and regularized random walks ranking for saliency detection," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1311 - 1322, 2016.

- [80] Y. Xu, J. Li, J. Chen, G. Shen and Y. Gao, "A novel approach for visual Saliency detection and segmentation based on objectness and top-down attention," in 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, 2017.
- [81] A. Azaza, J. v. d. Weijer, A. Douik and M. Masana, "Context proposals for saliency detection," *Computer Vision and Image Understanding*, 2018.
- [82] Y. Li, X. Hou, C. Koch, J. M. Rehg and A. L. Yuille, "The Secrets of Salient Object Segmentation," in *Computer Vision and Pattern Recognition*, Columbus, 2014.
- [83] E. E and E. A, "Visual saliency estimation by nonlinearly integrating features using region covariances," *J. Vision*, vol. 13, 2013.
- [84] S. Wang, M. Jiang, X. M. Duchesne, E. A. Laugeson, D. P. Kennedy, R. Adolphs and Q. Zhao, "Atypical visual saliency in autism spectrum disorder quantified through," *Neuron*, p. 604–616, 2015.
- [85] X. Hou, J. Harel and C. Koch, "Image Signature: Highlighting Sparse Salient Regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194-201, 2011.
- [86] B. W. Tatler, R. J. Baddeley and I. D. Gilchrist, "Visual correlates of fixation selection: effects of scale and time," *Vision Research*, vol. 45, no. 5, pp. 643-659, 2005.
- [87] A. Borji, M.-M. Cheng, H. Jiang and J. Li, "Salient object detection: A benchmark," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5706 - 5722, 2015.
- [88] X. Li, Y. Li, C. Shen, A. Dick and A. V. D. Hengel, "Contextual Hypergraph Modeling for Salient Object Detection," in *IEEE International Conference on Computer Vision*, Sydney, 2013.
- [89] K. A. Ehinger, B. Hidalgo-Sotelo, A. Torralba and A. Oliva, "Modeling Search for People in 900 Scenes: A combined source model of eye guidance," *Visual Cognition*, pp. 945-978, 2009.
- [90] T. Judd, K. Ehinger, F. Durand and A. Torralba, "Learning to predict where humans look," in *IEEE 12th International Conference on Computer Vision*, Kyoto, 2009.
- [91] W. Qiu, X. Gao and B. Han, "A superpixel-based CRF saliency detection approach," *Neurocomputing*, vol. 244, no. 28, pp. 19-32, 2017.
- [92] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong and S. J. Maybank, "Salient Object Detection via Structured Matrix Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 818-832, 2016.
- [93] L. Zhang et al., "Ranking Saliency", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 9, pp. 1892 - 1904, 2017.
- [94] W. Zhu, S. Liang, Y. Wei and J. Sun, "Saliency Optimization from Robust Background Detection," in *Computer Vision and Pattern Recognition*, Columbus, 2014.

[95] R. Margolin, A. Tal and L. Zelnik-Manor, "What Makes a Patch Distinct?," in *Computer Vision and Pattern Recognition*, Portland, 2013.

[96] J. Shi, Q. Yan, L. Xu and J. Jia, "Hierarchical Image Saliency Detection on Extended CSSD," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 717-729, 2015.