دادگان:

دادگان استفاده شده در این گزارش از دادگان ارائه شده توسط گروه تحقیقاتی آمریکایی پنس[[1]](#footnote-1) تهیه شده است. این دادگان شامل 15 دسته تصویر است که ورژن قدیم­تر آن (13 دسته­ای) توسط نویسنده مقاله مرجع مورد استفاده قرار گرفته است. این دادگان از لینک زیر دانلود شده است.

[http://www-cvr.ai.uiuc.edu/ponce\_grp](http://www-cvr.ai.uiuc.edu/ponce_grp/)

تصاویر همگی خاکستری و در سایز 256×256 است. از هر دسته داده 100 داده برای تعلیم و مابقی برای تست قرار داده شد است. برای استخراج کدبوک از 50 داده از دادگان تعلیم هر دسته استفاده می­شود. به منظور سهولت و همچنین سرعت بخشیدن به انجام آزمایش­ها از این مجموعه دادگان سه دسته داده (Simple2 و Simple1) استخراج شده است. در جدول زیر تعداد دادگان تعلیم و تست هر دسته معرفی شده است:

جدول 1: تعداد دادگان تعلیم و تست دادگان Simple1

|  |  |  |  |
| --- | --- | --- | --- |
| تعداد دادگان تعلیم | | تعداد دادگان تست | |
| جنگل | ساختمان بلند | جنگل | ساختمان بلند |
| 100 | 100 | 228 | 256 |

جدول 2: تعداد دادگان تعلیم و تست دادگان Simple2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| تعداد دادگان تعلیم | | | تعداد دادگان تست | | |
| جنگل | کوه | خیابان | جنگل | کوه | خیابان |
| 100 | 100 | 100 | 228 | 274 | 192 |

|  |
| --- |
| با توجه به متن مقاله برخی از دسته تصاویر به دلیل شباهت بیشتر قابلیت تفکیک کمتری دارد (مثل تصاویر حومه شهر و تصاویر شهری). از این رو در مقاله استفاده از یک ساختار سلسه­مراتبی در تصمیم­گیری پیشنهاد شده است. از آنجا که در این گزارش ساختار سلسه، روش استخراج ویژگی SIFT و ... استفاده نمی­شود، انتظار کارایی بالا از مدل نمی­رود. بنابراین دسته تصاویری که تمایز بیشتری دارند، انتخاب شده است تا تنها روش مدلسازی را بتوان پیاده­سازی کرد. |

استخراج کدبوک و ویژگی:

در مقاله چهار روش مختلف (Evenly Sampled Grid، Random Sampling و ...) برای تقطیع تصویر پیشنهاد شده است. در این گزارش تصویر به قطعات 11×11 بدون همپوشانی قطعه­بندی شده است و از این رو هر تصویر به 625 قطعه تقسیم شده است. همچنین هر قطعه تصویر در مقاله با یک روش استخراج ویژگی محلی چون روش استخراج ویژگی SIFT یا بردار 121 بعدی از مقادیر سطوح خاکستری تصویر نمایش داده شده است.

در این گزارش برای بدست آوردن کدبوک از هر دسته 50 داده تعلیم انتخاب و از هر تصویر 625 قطعه (مجموعه 93750) استخراج شد. بردار ویژگی هر قطعه مقادیر سطوح خاکستری قرار داده شده و با استفاده از روش kmeans به 200 خوشه دسته­بندی شد.

|  |
| --- |
| استخراج ویژگی محلی: در بسیاری از مقالات قدیمی تشخیص نوع صحنه، از ویژگی­های کلی چون هیستوگرام رنگ، توزیع فرکانسی و ... استفاده می­شد. در مقالات جدید مزیت ویژگی­های محلی اثبات شده و استخراج ویژگی از بخش­های کوچکی از تصویر انجام می­شود.  روش استخراج ویژگی SIFT: روش SIFTیک روش استخراج ویژگی محلی است که به تغییر سایز، تغییر مکانی و چرخش تصویر وابسته نیست.  روش kmenas: این روش یک روش خوشه­بندی است. حل آن یک مسئله NP-hard است ولی با الگوریتم­هایی کارا به مینیمم محلی همگرا می­­شود. در این روش دو مرحله تخصیص دادگان به خوشه و به روز کردن میانیگن خوشه­ها به طور تکراری انجام می­شود تا مقادیر خوشه­ها ثابت بماند. |

روش kmeans: برای پیاده­سازی روش kmeans از نرم­افزار متلب به صورت زیر استفاده شده است. در این تابع patch بردار 121×93750 بعدی و NumCodeWords برابر 200 است. کدورودهای کدبوک در متغیر Codewords قرار داده شده و در واقع میانگینی از دادگان قرار گرفته در هر دسته است.

[idx,Codewords] = kmeans(patch,NumCodeWords);

این روش قطعه­بندی و استخراج ویژگی برای سرعت در انجام آزمایش­ها اتخاذ شده ولی برای بهبود نتایج می­توان موارد زیر ار مورد بررسی قرار داد:

* سایز قطعات متغیر در نظر گرفته شود و به 11×11 نرمال شود.
* قطعات با همپوشانی در نظر گرفته شود.
* از ویژگی SIFT به جای سطوح خاکستری تصویر استفاده شود. در اکثر کدهای موجود SIFT بر کل تصویر اعمال شده و محل ویژگی­ها و ویژگی­های محلی استخراج می­شود. در این پروژه باید از هر قطعه یک بردار استخراج شود. از این رو فعلاً موفق به پیاده­سازی این روش نشدم.
* تاثیر تعداد کدبوک­ها بررسی شود.
* از هر کلاس کدبوک مجزا استخراج شود. سپس کدبوک­های کلاس­های مختلف تلفیق شده و کدورود­های مشترک حذف شود. در این صورت به نظر می­رسد ویژگی­های تمایزگرتری بدست آید. در دسته­بندی داکیومنت­های متنی نیز به نظر می­رسد که لغات مشترک در کلیه داکیومنت­ها جز کدبوک قرار داده نمی­شود.
* بردار سطوح خاکستری نرمالیزه شود. زیرا ممکن است تفاوت دو قطعه به دلیل تفاوت الگو نباشد بلکه تنها روشنایی تصویر تغییر کرده باشد.
* از سایر پیش­پردازش­های معمول تصویر برای نرمالیزه­کردن هیستوگرام و فیلترینگ استفاده شود.

در شکل زیر 200 عدد کدورد استخراج شده از 150 داده تعلیم از تصاویر (جنگل، خیابان و کوه) نمایش داده شده است:



شکل 1: کدورد­های استخراج شده از 150 داده تعلیم از تصاویر (جنگل، خیابان و کوه)

استخراج ویژگی محلی با توجه به کدبوک

اولین گام برای تعلیم مدل، استخراج ویژگی محلی با توجه به کدبوک است. بدین منظور ابتدا هر داده مشابه مرحله تعلیم به قطعات با اندازه 11×11 تقسیم بندی می­شود. بردار ویژگی معرف هر قطعه، بردار 121 بعدی مربوط به سطوح خاکستری است. این بردار با 200 بردار 121 بعدی کدبوک با استفاده از معیار فاصله کسینوسی مقایسه شده و کدبوک با بیشترین شباهت انتخاب می­شود. این کار برای کل 625 قطعه هر تصویر تکرار شده و تعداد هر یک از کدبوک­های برنده در یک بردار 200 بعدی ذخیره می­شود. بنابراین هر داده تصویر به یک بردار 200 بعدی تبدیل می­شود که هر مولفه آن تعداد کدورد مربوطه را در کل تصویرنشان می­دهد. در شکل زیر کدورود مربوط به یک تصویر جنگ و یک تصویر خیابان نمایش داده شده است:





شکل 2: شکل بالا مربوط به یک بردار ویژگی 200 بعدی از تصویر جنگل و شکل پایین مربوط به یک تصویر خیابان است.

|  |
| --- |
| فاصله کسینوسی:  برای اندازه­گیری شباهت دو بردار معیار­های مختلفی وجود دارد. در اینجا ما از معیار فاصله کسینوسی استفاده می­کنیم. برای دو بردار A و B با زاویه Ɵ فاصله کسینوسی به صورت زیر تعریف می­شود: |

مدل

مدل نایو بیز: مدل نایو بیز یک مدل دسته­بندی ساده است. این مدل با روش بیشینه­سازی درستنمایی و با فرم بسته قابل تعلیم است. در این مدل طبقه­بندی کننده، شرط اصلی استقلال ویژگی­ها به شرط دانستن متغیر پنهان (کلاس) است. گر چه این مدل قابلیت­ها و کاربردهای بسیاری دارد ولی به دلیل عدم امکان تعلق یک داده به چند دسته، از مدل­های دیگری چون LDA استفاده می­شود.

مدل LDA:

مدل مقاله Lei Fei Fei:

مدل MMNB: این مدل در واقع فرم کلی تری از مدل نایو بیز است. در این مدل با قرار دادن یک توزیع دریچله در بالا، قابلیت عضویت هر داده به چندین کلاس به مدل نایو بیز اضافه می­شود. این مدل در سال 2010 پیشنهاد شده [Shan 2010] و در آن از روش استنتاج تغییراتی[[2]](#footnote-2) استفاده شده است. خانم Shan علاوه بر روش استنتاج تغییراتی مشابه LDA یک روش استنتاج سریع را نیز پیشنهاد کرده است. کدها و برنامه­های مربوط به روش LDA، روش LDA تمایزی (DLDA)، روش MMNB و روش MMNB تمایزی (DMNB) با الگوریتم استنتاج تغییراتی سریع در سایت خانم Shan به آدرس زیر موجود است. الگوریتم استنتاج تغییراتی سریع پیشنهاد شده از پارامترهای کمتری استفاده می­کند در نتیجه مدل سریع­تر و با حجم حافظه کمتر و مستقل از حجم دادگان تعلیم می­یابد. بخشی از پیاده­سازی­های انجام شده در این گزارش توسط کدهای دانلود شده از آدرس زیر انجام شده است:

<http://www-users.cs.umn.edu/~shan/mmnb_code.html>

یادگیری مدل: در صورتی که تعلیم پارامترهای مدل با استفاده از روش معمول بیشینه­سازی امید ریاضی (EM) ممکن نباشد، از روش­های تقریبی استنتاج تغییراتی و نمونه­برداری گیبس استفاده می­شود.

استنتاج تغییراتی بیزی: استنباط بیزی از جمله روش­های رایج در یادگیری بیزی است که برای تقریب جواب با استفاده از یک سری فرض­های استقلال در توزیع پسین است. نکته مشکل ساز در یادگیری بیزی این است که برای حساب کردن توزیع پسین لازم است انتگرالی روی تمام حالات ممکن متغیرهای پنهان حساب شود. استنباط تغییراتی سعی در تقریب این انتگرال دشوار دارد تا یادگیری مدل و استنباط با آن را آسان­تر کند. به عبارتی دیگر:

1. روش استنباط بیزی تغییراتی تقریبی برای توزیع پسین می­دهد. با استفاده از این تقریب و داشتن پارامترهای مدل، می­توان [استنباط آماری](https://fa.wikipedia.org/wiki/%D8%A7%D8%B3%D8%AA%D9%86%D8%A8%D8%A7%D8%B7_%D8%A2%D9%85%D8%A7%D8%B1%DB%8C) روی داده­های دیده نشده انجام داد.
2. کرانی پایین برای درست نمایی حاشیه­ای روی داده­های آموزشی می­دهد. با استفاده از این کران می­توان پارامترهای مدل را یاد گرفت ایده­ی کلی این است که هرچه مقدار درست نمایی برای داده­های مورد نظر بیشتر باشد، پارامترها برای مدل و [داده­ها](https://fa.wikipedia.org/wiki/%D8%AF%D8%A7%D8%AF%D9%87_%D9%87%D8%A7) مناسب­تر هستند.

فرض استقلال بین توزیع­های پارامترهای مساله اساس روش استنتاج تغییراتی است. چنین فرضی باعث ایجاد خطا در نتیجه­ی نهایی خواهد شد، اما در قبال این خطا، سرعت بیشتری در یادگیری مدل به دست می آوریم.

Variational Message Passing (VMP):

|  |
| --- |
| توزیع بتا:  توزیع بتا، [توزیع احتمالی](https://fa.wikipedia.org/wiki/%D8%AA%D9%88%D8%B2%DB%8C%D8%B9_%D8%A7%D8%AD%D8%AA%D9%85%D8%A7%D9%84%DB%8C) پیوسته‌ای است که بر بازه ۰ تا ۱ تعربف می‌شود و دارای دو [پارامتر شکل](https://fa.wikipedia.org/wiki/%D9%BE%D8%A7%D8%B1%D8%A7%D9%85%D8%AA%D8%B1_%D8%B4%DA%A9%D9%84) است. این توزیع حالت خاصی از [توزیع دریکله](https://fa.wikipedia.org/wiki/%D8%AA%D9%88%D8%B2%DB%8C%D8%B9_%D8%AF%D8%B1%DB%8C%DA%A9%D9%84%D9%87) می‌باشد.    توزیع دریچله:  توزیع دیریکله یک توزیع پیوسته است. این توزیع بطور کلی حالت گسترش یافته [توزیع بتا](https://fa.wikipedia.org/wiki/%D8%AA%D9%88%D8%B2%DB%8C%D8%B9_%D8%A8%D8%AA%D8%A7) برای توابع چندمتغیره است. معمولاً از توزیع دیریکله به عنوان توزیع پیشین در [مدل سازی](https://fa.wikipedia.org/wiki/%D9%85%D8%AF%D9%84_%D8%B3%D8%A7%D8%B2%DB%8C) بیزی استفاده می­شود؛ چرا که توزیع دیریکله [مزدوج پیشین](https://fa.wikipedia.org/w/index.php?title=%D9%85%D8%B2%D8%AF%D9%88%D8%AC_%D9%BE%DB%8C%D8%B4%DB%8C%D9%86&action=edit&redlink=1&preload=%D8%A7%D9%84%DA%AF%D9%88%3A%D8%A7%DB%8C%D8%AC%D8%A7%D8%AF%2B%D9%85%D9%82%D8%A7%D9%84%D9%87%2F%D8%A7%D8%B3%D8%AA%D8%AE%D9%88%D8%A7%D9%86%E2%80%8C%D8%A8%D9%86%D8%AF%DB%8C&editintro=%D8%A7%D9%84%DA%AF%D9%88%3A%D8%A7%DB%8C%D8%AC%D8%A7%D8%AF%2B%D9%85%D9%82%D8%A7%D9%84%D9%87%2F%D8%A7%D8%AF%DB%8C%D8%AA%E2%80%8C%D9%86%D9%88%D8%AA%DB%8C%D8%B3&summary=%D8%A7%DB%8C%D8%AC%D8%A7%D8%AF%2B%DB%8C%DA%A9%2B%D9%85%D9%82%D8%A7%D9%84%D9%87%2B%D9%86%D9%88%2B%D8%A7%D8%B2%2B%D8%B7%D8%B1%DB%8C%D9%82%2B%D8%A7%DB%8C%D8%AC%D8%A7%D8%AF%DA%AF%D8%B1&nosummary=&prefix=&minor=&create=%D8%AF%D8%B1%D8%B3%D8%AA%2B%DA%A9%D8%B1%D8%AF%D9%86%2B%D9%85%D9%82%D8%A7%D9%84%D9%87%2B%D8%AC%D8%AF%DB%8C%D8%AF) (conjugate prior) برای [توزیع چندجمله ای](https://fa.wikipedia.org/wiki/%D8%AA%D9%88%D8%B2%DB%8C%D8%B9_%DA%86%D9%86%D8%AF%D8%AC%D9%85%D9%84%D9%87_%D8%A7%DB%8C) است.    توزیع چند جمله­ای (Multinomial):  توزیع چند تعمیم [توزیع دوجمله‌ای](https://fa.wikipedia.org/wiki/%D8%AA%D9%88%D8%B2%DB%8C%D8%B9_%D8%AF%D9%88%D8%AC%D9%85%D9%84%D9%87%E2%80%8C%D8%A7%DB%8C) است. در واقع در این توزیع به ازای n [آزمایش تصادفی](https://fa.wikipedia.org/wiki/%D8%A2%D8%B2%D9%85%D8%A7%DB%8C%D8%B4_%D8%AA%D8%B5%D8%A7%D8%AF%D9%81%DB%8C) و مستقل، k نتیجه هرکدام با احتمال بروز مشخص ثابت بروز می‌کنند. در واقع توزیع چند جمله‌ای احتمال بروز هرگونه ترکیبی از n برآمد تصادفی مستقل (که هرکدام می‌توانند از میان یکی از k برآمد ممکن باشند) را بدست می‌دهد. |

|  |
| --- |
| تابع gamma :  تابع گاما تعمیم تابع [فاکتوریل](https://fa.wikipedia.org/wiki/%D9%81%D8%A7%DA%A9%D8%AA%D9%88%D8%B1%DB%8C%D9%84) است از مجموعه [اعداد طبیعی](https://fa.wikipedia.org/wiki/%D8%A7%D8%B9%D8%AF%D8%A7%D8%AF_%D8%B7%D8%A8%DB%8C%D8%B9%DB%8C) به مجموعه [اعداد حقیقی](https://fa.wikipedia.org/wiki/%D8%A7%D8%B9%D8%AF%D8%A7%D8%AF_%D8%AD%D9%82%DB%8C%D9%82%DB%8C) و [مختلط](https://fa.wikipedia.org/wiki/%D8%A7%D8%B9%D8%AF%D8%A7%D8%AF_%D9%85%D8%AE%D8%AA%D9%84%D8%B7) و برای یک عدد مختلط با بخش حقیقی مثبت به شکل زیر تعریف می‌شود:    تابع digamma: |

|  |
| --- |
| فاصله Kullback-Leibler |

|  |
| --- |
| Jensen’s inequality  این نامساوی رابطه بین تابع محدب اعمال شده بر یک انتگرال و انتگرال یک تابع محدب را بیان می­کند. در ساده­ترین فرم، نامساوی جنسون بیان می­کند که تابع محدب اعمال شده بر میانگین از میانگین تابع محدب کمتر است. در تئوری احتمال نیز اگر x متغیر تصادفی و phi تابع محدب باشد، داریم: |

1. Ponce research group [↑](#footnote-ref-1)
2. Variational Inference [↑](#footnote-ref-2)