معرفي پیشین فرآیند دیرکله در چارچوب مدل‌های بی‌نای‌پارامتری

عاطقه جاویدی آل‌سعیدی، سمیه راهی‌ماهی کازرونی، مجید جعفری خالدی

چکیده:
مدل‌های آماری برای شناخت فرآیندی که داده‌ها از آن تولید شده، استفاده می‌شود. در پیش‌بینی مدل‌های فرض می‌شود متغیرهای تصادفی $Y_1, ..., Y_n$ نمونه‌ای تصادفی از توزیع $F$ هستند که متعلق به یک کلاس از خانواده توزیع‌های پارامتری است. اما در عمل همیشه نمی‌توان انتظار داشت که یک مدل پارامتری برای توصیف داده‌های مشابه باشد. در این شرایط می‌توان مدل‌های پارامتری را کل گذاشت و از مدل‌های انعطاف‌پذیر و نیرومندتر برای تحلیل داده‌های استفاده کرد. در چارچوب روش بی‌نای‌پارامتری با تعریف یک توزیع پیشین روی فضای کل توزیع‌های احتمالی و در نظر گرفتن آن برای توزیع متغیر تصادفی این انعطاف‌پذیری حاصل می‌شود.

بی‌نای‌پارامتری مورد استفاده قرار می‌گیرد. در این مقاله این فرآیند و خواص آن معرفی می‌شود.

واژه‌های کلیدی: توزیع دیرکله، فرآیند دیرکله، بی‌نای‌پارامتری.

1. مقدمه

تعریف 1.1 فضای خطی $S$ متناهی بعد است $k$ فضای متناهی بعد است $m$ عناصر در فضای باشد. در واقع $S$ یک تعریف $m$ عناصر خطی متناهی است، باشد. بعد یک فضای خطی متناهی بعد به صورت تعداد عناصر مستقل خطی که در فضای کل گسترده شده است، تعريف می‌شود. یک هدف مهم در مدل‌های پارامتری، برآورد یک مقدار قابل قبول برای پارامتر مدل، یعنی $\theta$ است. همان‌طور که اشاره شد در مدل‌های پارامتری فرض محدودیت‌های معنی‌دار از فرم توزیع ممکن است نتایج استنباط‌آمیز را نامعتبر سازد. اما می‌توان با کوار گذاشتن مفروضات پارامتری مدلی انعطاف‌پذیر و نیرومندتر ارائه کرد. این گونه مدل‌ها به نظر گرفته می‌شود که در آن تابع چگالی در فضای بزرگتر و پارامترها دارای بعد نامتناهی هستند.

فرض کنید داده‌ها تحت حکمی از متغیرهای تصادفی $Y_1, ..., Y_n$ معمولی توزیع می‌شود که $Y_1, ..., Y_n$ به طور مستقل از توزیع $B$ احتمالی و $Z$ مجموعه‌ای اندام‌بندی از فضای نمونهای $A$ و و zad $\mu(\cdot)$ تابع چگالی مناظر با اندام‌بندی $\mu(B) = P(Y \in B)$ احتمالی $P(A)$ و ارائه $\mu(\cdot)$ همان‌طور که این توزیع نامعلوم $f(x)$ است. مدل‌های پارامتری $\mu(\cdot)$ به صورت $f_0(x)$ بوده و در واقع یک عضو از خانواده توزیع‌های تعریف می‌شود که در آن $\mu$ که در آن $p \leq \mu$ هستند است. 

لازم به ذکر است که در این حالت فضای پارامتری یک زیرمجموعه از فضای خطی متناهی بعد است.

$A$ معلامه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس
$B$ معلامه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس

$\mu(\cdot)$ فضای هیئت علمی گروه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس

---

$^4$ Parametric Models

$^5$ Finite Dimensional Space
پیشین $\pi(G)$ و $f(x_i)$ یک تابع پارامتری در نظر گرفته می‌شود.

2. مدل طبق میانگین نابارماری: در این مدل طبق میانگین دارای توزیع تصادفی در نظر گرفته می‌شود و پیشین برای این توزیع انتخاب می‌شود. به عبارت دیگر $f \sim \pi(f)$

که یک توزیع پیشین است. پیشین مرسمون برای $f$ فراپارامتری گاویست (مولر, 2012).

3. رگرسیون نابارماری کامل: در این مدل هم تابع میانگین و هم توزیع خطا تصادفی متغیر $x$ و برای آنها توزیع پیشینی در نظر گرفته می‌شود.

برای دیدن جزئیات و مثال‌های بیشتر پیرامون روشهای پیشین و نابارماری، مراجعه کردن شیوه‌های مختلفی برای تعريف پیشین برای توزیع تصادفی وجود دارد (دی و همکاران، 1998). رایج‌ترین و پرکاربردترین آنها برمی‌نینی فراپارامتری تصادفی است. فراپارامتری چندمتغیری توزیع پیشینی تصادفی و میانگین توزیعی می‌باشد. به دلیل پیچیدگی محاسبات در روش نابارماری انتخاب نوع فراپارامتری تصادفی یکی در انتخاب باید باید صورت گیرد. به دنبال انتخاب شد که انتخاب $G$.

در واقع داریم:

$Y_1,\ldots, Y_n \mid G \overset{iid}{\sim} G$

$G \sim \pi(G)$.

در حالیکه اکثر مدل‌های آماری مختلف بر اساس روش پیشین نابارماری مورد بررسی قرار گرفته‌اند. در اینجا فرض نابارماری در یک مدل در چند متغیره بایان می‌شود. به عبارت دیگر یک پیشین مزدوج باشد. از جمله این فرض‌ها، فراپارامتری است. فرگوسن (1973) فراپارامتری را به عنوان یک توزیع پیشین روی فضای تمام توابع توزیع تعریف کرده، یک پیشین ناگهانی که موجب راه اندازی محاسبات و همچنین استفاده دقیق و کارا می‌شود. و کستروش روش‌های محاسباتی بیزی مانند روش چهار روش مونت کارل، توزیع پیشین $\pi(f(x_i)) + \varepsilon_i$ و متغیرهای وابسته $y_i$ را به صورت $\varepsilon \sim p_\epsilon(\alpha)$ که $\alpha$ از نظر سایر بگیری. در مدل رگرسیون بیزی نابارماری، توزیع خطا $\varepsilon$ و $y_i, x_i$ نابارماری در نظر گرفته می‌شود. اما در مدل‌های رگرسیون بیزی نابارماری آنها نابارماری در نظر $G$ که $\alpha_i \sim G$ است، می‌تواند به سه حالت زیر یکی‌شود:

1. مدل خطا نابارماری: در این مدل فرض می‌شود $\alpha_i \sim G$ دارای توزیع تصادفی $G$ است، می‌تواند به سه حالت زیر یکی‌شود:

$\varepsilon_i \sim G$.

Infinite Dimensional Space
2 توزیع دیریکل

توزیع دیریکل در نظریه احتمال و آمار یک توزیع پیوسته است.

در واقع این توزیع یک توزیع چند پارامتری تعمیمی‌افته از توزیع
بتا می‌باشد. معمولاً از توزیع دیریکل به عنوان توزیع پیشین
در استنباط به‌یزی استفاده می‌شود، چراکه این توزیع یک پیشین
مزدوج برای پارامترهای توزیع چند جمله‌ای و توزیع راسته‌ای
است.

تعریف 2.1 فرض کنید $\Theta = \{\theta_1, \ldots, \theta_n\}$ یک توزیع احتمال
روی فضای گستن $X = \{x_1, \ldots, x_n\}$ متغیر تصادفی روی $X$
باشد به طوری که $p(X = x_i) = \theta_i$.

توزیع دیریکل روی $\Theta$ به صورت

$$p(\Theta | \beta) = \frac{\Gamma(\beta_0)}{\prod_{i=1}^{n} \Gamma(\beta_i)} \prod_{i=1}^{n} \theta_i^{\beta_i - 1}$$

می‌باشد، که در آن $0 < \beta_i > 0$ و $\sum_{i=1}^{n} \theta_i = 1$, $\theta_i > 0$.

گشتاورهای این توزیع عبارت است از

$$E(\theta_i) = \frac{\beta_i}{\beta_0},$$

$$Var(\theta_i) = \frac{\beta_i (\beta_0 - \beta_i)}{\beta_0^2 (\beta_0 + 1)},$$

$$Cov(\theta_i, \theta_j) = \frac{-\beta_i \beta_j}{\beta_0^2 (\beta_0 + 1)}.$$
۳ فرآیند دیریکله

فرآیند دیریکله یک توزیع از توزیع دیریکله است که می‌تواند به‌عنوان یک توزیع پیشین برای میزان انتخاب هر یک از فضاهای باشند.

نتیجه‌گیری از توزیع دیریکله

یک روش تولید نمونه از توزیع دیریکله استفاده از طرف نیز است. فرض کنید می‌خواهیم از توزیع دیریکله $\Theta \sim Dir(\alpha M, X_1, \ldots, X_n)$ نمونه تولید کنیم. برای شروع فرض کنید $\alpha$ یک گویی با رنگ‌های مختلف در طرف وجود دارد، به‌طوریکه $\alpha$ یک از رنگ اول، $am_1$ یک از رنگ دوم و ... هر رنگ در فضای $B_1, \ldots, B_k$ خواهد بود.

شکل ۲ مراحل تولید داده از یک توزیع دیریکله بر اساس طرف پولیا.
وجود فرا인دهای دیریکله به عنوان یک فرایند تصادفی تعريف شده روی فضای \((\chi, \beta)\) از طریق قضیه سارگاری کلمنگروف (کلمگروف، 1963) توسط فرگوسن اثبات گردید. بلکه (1963) اثبات دیگری را ارائه داده است.

به عنوان یک فرایند دیریکله توزیعی روی انتظاره احتمال

\(G(B) \in \beta\) است، بایر هر توزیع \(G\) تعريفی است.

با توجه به تعريف اين فرايند، مي توان گفت:

\[
G(B) \sim Beta\{aG_0(B), \alpha(1 - G_0(B))\}
\]

\[
E(G(B)) = \frac{G_0(B)}{1 + \alpha}
\]

\[
Var(G(B)) = \frac{G_0(B)(1 - G_0(B))}{1 + \alpha}
\]

شك Highway برای \(100\) داده اول بوسیله \(DP(\alpha G_0)\) با توزیع مرکزی نرمال استاندارد و به ترتیب 1000، 5000، 10000 نمودار داده شده است.
لازم به ذکر است که تکیه‌گاه‌های توزیع متغیرهای مستقل و هم‌توزیع از $V_1, V_2$ نیاز به مقدار می‌دارد.

$\beta, \alpha, G_0$ مقدار توزیع $\beta(\alpha, G_0)$ است. بنابراین با توجه به احتمال $\frac{\beta(\alpha, G_0)}{\alpha + n}$ و منظور به فرض $\beta(\alpha, G_0) = 1$, $\alpha = 0$, $G_0 = V_1$، $\beta(\alpha, G_0) = \frac{\alpha}{\alpha + n}$ و $G_0 = \sum_{i=1}^{\alpha} W_i$, احتمال $\beta(\alpha, G_0)$ مشابه در هر $\alpha$ وجود دارد.

فرض کنید $X_1, X_2, \ldots$ متغیرهای مستقل و هم‌توزیع با $\beta(1, \alpha)$، آنگاه $G_0$ مجموعه متغیرهایی باشد که به ترتیب: 

$x = 1, \ldots, X_1, \ldots, X_n$.

شناخته شده فرآیند دیریکله با اساسی پیش‌بینی می‌شود.

تکیه‌گاه این فرضیه که در محاسبات بیزی کاربردی به صورت زیر است:

$G \mid X_1, \ldots, X_n \sim DP(\alpha G_0 + \sum_{i=1}^{n} \delta_{X_i})$

تعریف $G \mid X_1, \ldots, X_n \sim DP(\alpha G_0 + \sum_{i=1}^{n} \delta_{X_i})$ این فرض کنید $G \sim DP(\alpha G_0)$ باشد.

$G \mid \alpha, G_0 \sim DP(\alpha G_0)$}

**قضیه 3.3 (فرمول 1973)**

فرض کنید $X_1, \ldots, X_n \mid G \sim \beta(\alpha, G_0)$ باشد.

** قضیه 3.3 (فرمول 1973)**

فرض کنید $X_1, \ldots, X_n \mid G \sim \beta(\alpha, G_0)$ باشد.
فراآیند دیریکلت آمیخته با هسته نرمال میانگین بر پارامتر مناسبی، 
در دامین

\[ y_i | G, \sigma^2 \sim \int N(y_i | \mu, \sigma^2)G(d\mu) \]

شکل 4. فراآیند دیریکلت آمیخته. پیچشی از اندوزه احتمال تصادفی به صورت سلسله مراتبی

\[ y_i | \theta_i \sim p(y_i | \theta_i), \]

\[ (\theta_1, \ldots, \theta_n) \sim p(\theta_1, \ldots, \theta_n), \]

\[ G \sim DP(\alpha G_0) \]

شکل 4: یک مدل آمیخته با هسته نرمال و G

\[ G | \eta \sim DP(\alpha \eta G_0) \]

\[ \eta \sim \pi(\eta) \]

آمیخته از فراآیند دیریکلت با پارامتر ذبیحه G دقت توزیع \( \alpha \eta \) آمیخته از فراآیند دیریکلت با پارامتر ذبیحه G 

\[ G | \eta \sim \int \pi(\eta)DP(\alpha \eta G_0)d\eta \]

2.4 فراآیند دیریکلت آمیخته

اگرچه انتخاب فراآیند دیریکلت به عنوان پیشین یک توزیع نامعلوم، مناسب بوده و ویژگی‌های جالبی به همراه دارد، اما هر تحقیق

از این فراآیند یک توزیع گسترش انتخاب باین و توزیع پیشین نامعلوم کمی داشته بوده است موجب بنظر

نیست. یکی از روش‌هایی که برای برطرف کردن این موضوع استفاده می‌گردد. مدل فراآیند دیریکلت مورد استفاده قرار می‌گیرد. مدل فراآیند دیریکلت

آنها از توزیع نامعلوم، G نیز بر ماهیت خوشبندی باینوسی. مدل سلسله مراتبی (79) نیز بر پیشین گسترش احتمال انتخاب می‌شنود.

\[ G | y \sim F(y) \]

\[ \bar{F} \sim \frac{DP(\alpha G_0)}{G} \]

\[ y_i | G, \theta_i \sim p(y_i | \theta_i), \]

\[ (\theta_1, \ldots, \theta_n) \sim p(\theta_1, \ldots, \theta_n), \]

\[ G \sim DP(\alpha G_0) \]

4 مدل‌های مبتنی بر فراآیند دیریکلت

1.4 آمیخته از فراآیند دیریکلت

(1974) معرفی گردید. مدل‌های سلسله مراتبی که در آن یک سطح شامل فراآیند دیریکلت است کاربرد دارد. بطور کلی اگر توزیع پایه G و پارامتر ذبیحه α تصادفی باشد، عیت

\[ G | \eta \sim DP(\alpha \eta G_0) \]

\[ \eta \sim \pi(\eta) \]

انجامدها از فراآیند دیریکلت با پارامتر ذبیحه G دقت توزیع \( \alpha \eta \) انتخاب باشد. در این حالت می‌نویسیم

\[ G | \eta \sim \int \pi(\eta)DP(\alpha \eta G_0)d\eta \]

16 Mixture of Dirichlet Processes

17 Dirichlet Process Mixture Model

18 Convolution
خوشه و نمونه‌های متغیر از خوشه‌های متغیر با 
مجموعه مقداری از خوشه‌های متغیر با 
همچنین 

\[ \theta_1, \ldots, \theta_n \sim G_0, \]

\[ p(\xi_1, \ldots, \xi_n | \alpha, n^*) = \frac{\Gamma(\alpha)}{\Gamma(n^*)} \frac{\alpha^n}{\Gamma(n^*)} \prod_{j=1}^{n^*} \Gamma(n_j), \]

که در آن 

\[ n_j = \sum I(\xi_j = j), \]

به دلیل خارج کردن توزیع نامتناهی بعدی 

\[ G \]

وعنداین 

\[ y_1 | W_h, \theta_h \sim \sum_{h=1}^{\infty} W_h p(y_1 | \theta_h) \]

به طور گسترده مورد استفاده قرار گیرد.

مدل (94) را می‌توان با استفاده از ساختار معرفی شده بطور یک برابری 

\[ G \]

در فرآیند دیریکلته به صورت زیر تعیین کرد:

\[ \theta_h \sim G_0 \]

\[ W_h = V_h \prod_{k<h} (1 - V_h) \]

\[ \alpha \sim Beta(1, \alpha) \]

\[ G = \sum_{h=1}^{\infty} W_h \theta_h \]

\[ \alpha \sim \Gamma(n^* + \alpha) \frac{\alpha^n}{\Gamma(n^*)} \prod_{j=1}^{n^*} \Gamma(n_j) \]

\[ y_1 | \theta_h \sim \sum_{h=1}^{\infty} W_h \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]

\[ \sum_{h=1}^{\infty} W_h \theta_h \]

\[ G \]
در واقع هنگامی که

\[ \epsilon = 0 \to \epsilon \to 0 \]

به فرآیند دیریکله همگرا می‌شود. به عبارت دیگر با در نظر گرفتن

یک مقدار 

\( \epsilon > 0 \) مناسب، می‌توان تقریبی مناسب از فرآیند دیریکله در اختیار داشت. لازم به ذکر است ایشواران و جیمز (۲۰۰۱) تعريف تذگری از فرآیند دیریکله بوسیله شده ارائه کردند. در این

تعريف تقریبی دریگر از فرآیند دیریکله به صورت

\[
\sum_{i=1}^{\infty} W_i \delta_{\theta_i} \approx \sum_{i=1}^{N} W_i \delta_{\theta_i}
\]

به عنوان فرآیند دیریکله بریده شده معرفی شد و نشان دادند

\[ W_N = 1 - \sum_{i=1}^{N-1} W_i \]

بوده و پارامتر

\( N \) تعداد نقاط جرم مورد استفاده را مشخص می‌کند.

همچنین مقدار رابطه مستقیم با

\( N \) اکثریت کرده که کنترل کننده خوشودها در بین واحدهای آزمایشی است. هنگامی که

نکته که نکته که

کوشیده باشد

محاسبات ساده است اما تقریبی مورد نظر دقیق نیست. رویکردی که برای انتخاب

\( N \) بررسی‌های انتخاب شده مقدار مورد انتظار

به‌گونه‌ای که

\[ E(W_N) \approx \epsilon \]

باشد به عبارت دیگر

ایشواران و جیمز (۲۰۰۱) نشان دادند در این صورت داریم

\[ N \approx 1 + \frac{\log \epsilon}{\log(1 + \alpha)} \]
فرض می‌شود توزیع پیشین متغیر پهن و نرم است و بدين ترتيب
مدل حاصل بروریت ۱۹ نامه‌ی می‌شود. این در حالی است که چنین
فرضی ممکن است برقرار نباشد. برای حل این مسئله رهیافت‌های
متغیران وجود دارد (کافیو و همکاران، ۲۰۰۷). یک رهیافت بر
منای اتخاذ رهیافت بیز نیمه‌پارامتری ۱۰ است که در آن پارامترهای
مدل متعارف پهن به دو دسته تقسیم شده و برای برخی از آنها
روش پارامتری می‌باشد و برای بقیه پیش‌بینی نیمه‌پارامتری برابر
اساس فرآیند دیرکلکه توزیع می‌شود. در صورتی که یک مدل بروریت
عمیق‌احطه به توزیع پهن از فرآیند دیرکلکه آمیخته ۱۱ شکل می‌گیرد (جراح و
همکاران، ۲۰۰۷). در مدل‌های آمیخته خطي توزیع‌یافته ۱۲ برای
مدل‌های متغیر تصادفی اغلب از توزیع نرمال استفاده می‌شود. اما
ممکن است یک توزیع برقرار نباشد. از این رو استفاده از پیشین
فرآیند دیرکلکه برای توزیع نامعلوم اثرات تصادفی متغیر به استنباط
دقت‌تر و کارایی بیشتر مدل می‌شود. کونک و همکاران (۲۰۱۱)
برای مدل‌های متغیر تصادفی از فرآیند دیرکلکه استفاده کرده‌اند.

بحث و نتایج گیری

در این مقاله فرآیند دیرکلکه و ویژگی‌های مهم آن به عنوان یک
پیشین مناسب در مدل‌های پیش‌بینی معرفی شد. همان‌طور
که بیان شد این فرآیند ضمن اینکه به عنوان یک پیشین مزدوج
در تسریع محاسبات بیزی نقش مهمی دارد و ویژگی خوش‌مندی 
داده‌ها ناهمگنی بین آنها را کنترل کرده و این امر منجر به بهبود
و کارایی مدل می‌شود. این خواص DP در سال‌های اخیر موجب
محبوبیت آن در گستره‌های وسیعی از مسائل پیش‌بینی مانند پردازش
تصادف و ایدئولوژی شده است. یکی از کاربردهای آن همکاری متغیر
به عنوان تحقیقات آینده می‌باشد. نظر داشته استفاده از این فرآیند در
مدل‌های پیش‌بینی نارانگی و پیش‌بینی فضایی است.

کرد. مدل سلسله مراتبی

\[ Y_i|b_i, \eta \sim f(y_i|b_i, \eta), \]

\[ b_i|G \sim G, \]

\[ G \sim DP(\alpha, G_0) \]  

(۹)

(۱۰)

پیروی می‌کند. راپه‌تی (۴۴) در واقع پیانگر این است که اثرات
تصادفی با احتمال مشابه یکی از مقادیر قبیل یا و با
احتمال \( \frac{1}{n+1} \) یک مقدار حذفی از را می‌گیرد. همان‌طور
که گفته استفاده از این ساختار به دلیل عدم توزیع با
دید انتخابی، \( G \) سادگی همبستگی را به دنبال دارد. اما مفهوم
خوشه‌بندی به عنوان یک ویژگی مهم DP با شکست نمی‌دهد.
بنابراین مشابه آنچه که در بخش‌های قبیل پیانگر با استفاده از
ویژگی خوشه‌بندی \( DP \) می‌توان یک مدل ساده‌تر و کارایی‌تر را ارائه
داد. یکی از کاربردهای مدل متعارف پهن در تحلیل داده‌های دو
ویژگی است. در این چارچوب، داده‌های دو-ویژگی را به یک متعارف
پهن پوسته ارتباط می‌دهند به این صورت که با استفاده از یک
مقدار آستانه متغیر پوسته در دو رده کاملاً شده و متغیر باشگ
معرفی می‌شود. به‌طور کلی در اغلب موارد، تحقیقات بحل آمده

14 Probit Model
15 Bayesian Semiparametric Model
16 Dirichlet Process Mixture
17 Generalized Linear Mixed Models
مراجع


