دانشگاه صنعتی امیرکبیر

(پلی تکنیک تهران)

گزارش سمینار درس بهینه‌سازی محدب

نگارش

امین عبدی‌پوراصل

۴۰۱۱۳۳۰۱۱

استاد درس

دکتر امیرمزلقانی

بهمن ۱۴۰۲

فهرست مطالب

صفحه

فهرست اشکال

صفحه

۱- مقدمه: بیان مفاهیم (خارج از مقاله)

۱-۱- رگرسیون تصویر به تصویر[1]

رگرسیون تصویر به تصویر یک نوع کار بینایی کامپیوتری است که هدف آن پیش‌بینی خروجی پیوسته برای هر پیکسل در یک تصویر ورودی است. برخلاف کارهای طبقه‌بندی که هدف آن اختصاص یک برچسب به کل تصویر یا هر پیکسل است، وظایف رگرسیون شامل پیش‌بینی یک مقدار عددی برای هر پیکسل یا مجموعه‌ای از نقاط هدف در تصویر است.

در زمینه رگرسیون تصویر به تصویر، ورودی معمولا یک تصویر است و خروجی یک تصویر متناظر است که در آن هر پیکسل حاوی یک مقدار پیوسته به جای یک برچسب گسسته است. این را می‌توان در برنامه‌های مختلف مانند بهبود تصویر، رنگ بندی، تخمین عمق و غیره استفاده کرد.

رگرسیون تصویر به تصویر شامل یادگیری یک نقشه برداری از تصاویر ورودی به تصاویر خروجی با ارزش پیوسته است که به مدل اجازه می‌دهد تا پیش بینی‌های دقیقی برای هر پیکسل در تصویر انجام دهد.

۱-۲- مدل‌های انتشار[2]

مدل‌های انتشار، به‌ویژه در زمینه مدل‌سازی مولد، به دسته‌ای از مدل‌های احتمالی اشاره می‌کنند که برای تولید نمونه‌های واقعی از توزیع داده‌های معین استفاده می‌شوند. این مدل‌ها به عنوان مدل‌های مولد مبتنی بر امتیاز[3]

[1] Image-to-image regression

[2] Diffusion models

[3] Score-based generative models

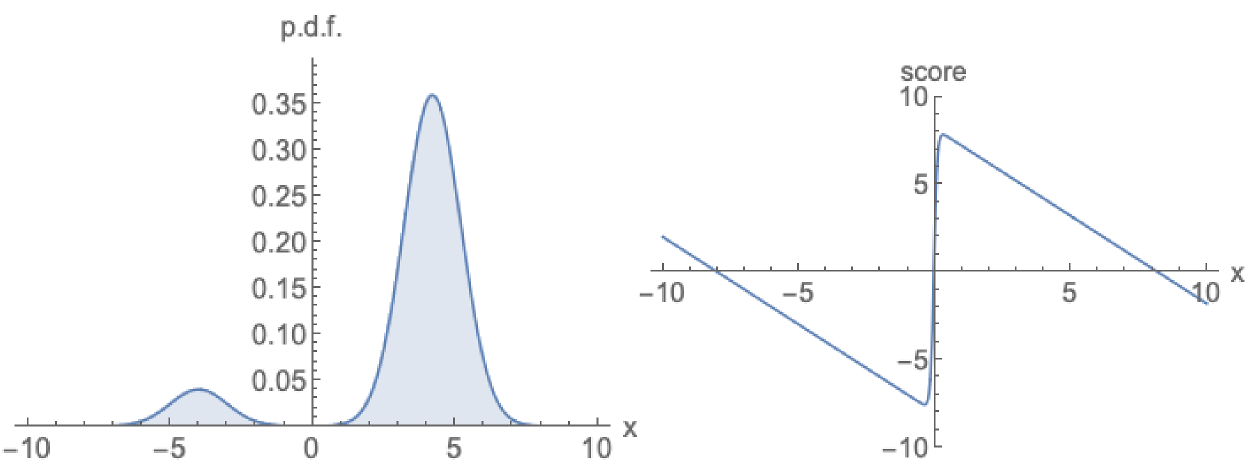
نیز شناخته می‌شوند. ایده پشت مدل‌های انتشار مدل‌سازی فرآیند تولید داده به عنوان یک سری مراحل تکراری است که در آن نویز به تدریج به یک نقطه داده اولیه اضافه می‌شود و منجر به تولید یک نمونه می‌شود.

در اینجا یک نمای کلی از نحوه عملکرد مدل‌های انتشار آورده شده است:

۱. فرآیند انتشار: مفهوم کلیدی فرآیند انتشار است که پخش یا انتشار نویز را در یک تصویر شبیه سازی می‌کند. این فرآیند با یک نقطه داده تمیز (به عنوان مثال، یک تصویر) شروع می‌شود و به طور مکرر مراحل انتشار را اعمال می‌کند و در هر مرحله نویز ایجاد می‌کند.

۲. مدل سازی مبتنی بر امتیاز: رویکرد مبتنی بر امتیاز شامل یادگیری یک تابع امتیاز است که شیب احتمال ورود به سیستم را با توجه به داده‌ها نشان می‌دهد. این تابع امتیاز، فرآیند انتشار را هدایت می‌کند و به تولید نمونه‌های واقعی کمک می‌کند. مدل‌سازی مولد مبتنی بر امتیاز بر یادگیری توزیع داده‌ها به طور غیرمستقیم با یادگیری گرادیان لگاریتم احتمال تمرکز دارد. بدین صورت:

$$\nabla_x \log p(x)$$

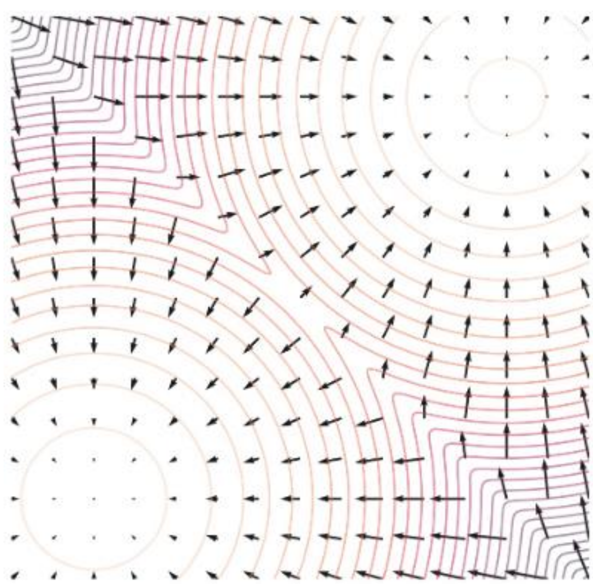


شکل ۱: نمایشی از در آوردن امتیاز از روی تابع چگالی احتمال

۳. فرآیند مولد: در طول تولید، مدل با یک توزیع ساده (مثلاً نویز گوسی) شروع می‌شود و از تابع امتیاز آموخته شده برای هدایت فرآیند انتشار استفاده می‌کند و به تدریج نویز را به نمونه ای شبیه به توزیع داده واقعی تبدیل می‌کند.

۴. آموزش: مدل با بهینه سازی پارامترها برای به حداقل رساندن شکاف بین نمونه‌های تولید شده و توزیع واقعی داده‌ها آموزش داده می‌شود. آموزش شامل به حداکثر رساندن احتمال داده‌های مشاهده شده تحت مدل است.

مدل‌های انتشار در تولید نمونه‌های با کیفیت بالا امیدوارکننده بوده و در حوزه‌های مختلف از جمله تولید تصویر و سنتز داده‌ها به کار گرفته شده‌اند. آنها جایگزینی برای مدل‌های مولد سنتی مانند شبکه‌های متخاصم مولد[۴] (GAN) و رمزگذارهای خودکار متغیر[۵] (VAE) ارائه می‌دهند. مدل‌های مبتنی بر امتیاز نیازی به داشتن یک ثابت نرمال‌کننده قابل حمل ندارند، و می‌توانند مستقیماً با تطبیق امتیاز یاد بگیرند. علاوه بر این، مدل‌های مبتنی بر امتیاز با مدل‌های جریان عادی ارتباط دارند، بنابراین امکان محاسبه احتمال دقیق و یادگیری نمایش را فراهم می‌کنند. علاوه بر این، مدل‌سازی و تخمین امتیازها، حل معکوس مسئله را با کاربردهایی مانند نقاشی درون تصویر، رنگ‌آمیزی تصویر، سنجش فشاری و بازسازی تصویر پزشکی (مانند CT، MRI) تسهیل می‌کند. شکل ۲، مثالی از توابع امتیاز به همراه کانتورهای تابع را به ما نشان می‌دهد.



شکل ۲: تابع امتیاز (میدان برداری) و تابع چگالی (خطوط) احتمال مخلوطی از دو گاوسی

[4] Generative Adversarial Networks

[5] Variational Autoencoder

۱-۳- مجموعه‌های پیش‌بینی کنترل ریسک[6] (RCPS)

مجموعه‌های پیش‌بینی کنترل ریسک یک روش آماری است که در یادگیری ماشین برای انجام پیش‌بینی‌هایی با ضمانت‌های ریسک قابل اندازه‌گیری استفاده می‌شود. این روش به ویژه در تنظیماتی که ارائه پیش‌بینی‌های قابل اعتماد با فواصل اطمینان مرتبط یا محدودیت‌های ریسک ضروری است مفید است، مانند مراقبت‌های بهداشتی، مالی، یا فرآیندهای تصمیم‌گیری حیاتی.

نمای کلی از روش RCPS:

۱. تولید مجموعه پیش بینی: رویه RCPS مجموعه ای از پیش بینی‌ها را به جای یک برآورد واحد تولید می‌کند. هدف این مجموعه پیش بینی‌ها پوشش دادن مقدار واقعی متغیر هدف با احتمال مشخصی است.

۲. کنترل ریسک: RCPS بر کنترل ریسک مرتبط با پیش بینی‌های خود تمرکز می‌کند. ریسک به این احتمال اشاره دارد که ارزش واقعی خارج از مجموعه پیش بینی قرار می‌گیرد. با کنترل این ریسک، رویه تضمین‌های قابل اعتمادی را در مورد کیفیت پیش بینی‌های خود ارائه می‌دهد.

۳. ضمانت‌های قابل اندازه گیری: RCPS تضمین‌های قابل اندازه گیری در مورد خطر مرتبط با پیش بینی‌های خود ارائه می‌دهد. به عنوان مثال، ممکن است تضمین کند که با احتمال ۹۵ درصد، مقدار واقعی در مجموعه پیش بینی قرار می‌گیرد. این به کاربران اجازه می‌دهد تا قابلیت اطمینان پیش‌بینی‌ها را ارزیابی کنند و بر اساس تحمل ریسک تصمیمات آگاهانه بگیرند.

۴. سازگاری: RCPS را می‌توان با وظایف و تنظیمات مختلف پیش بینی تطبیق داد. این می‌تواند انواع مختلفی از مدل‌های پیش بینی و توابع خطا را در خود جای دهد و آن را برای کاربردهای مختلف انعطاف پذیر می‌کند.

[6] Risk-Controlling Prediction Sets

۲- مفاهیم مقاله

۲-۱- هدف مقاله

اخیراً، کارهای اســـاســـی بر روی نمونه‌برداری از طریق تطبیق امتیاز و مدل‌های انتشـــار نویز زدایی، راه را برای کلاس جدیدی از مدل‌های مولد مبتنی بر امتیاز، که معادله دیفرانســیل تصـادفی زمان معکوس[۷] (SDE) را حل می‌کنند، هموار کرده است. این مدل‌ها به‌طور قابل‌توجهی در تولید نمونه بدون قید و شرط (یعنی شروع از نویز تصــادفی) و شــرطی (مانند رنگ‌آمیزی، حذف نویز، وضــوح فوق‌العاده یا طبقه‌بندی) در زمینه‌های مختلف مؤثر بوده‌اند. به عنوان مثال، مدل‌های مولد مبتنی بر امتیاز برای مشـــکلات معکوس در بینایی رایانه‌ای عمومی‌و تصویربرداری پزشکی، تولید شکل‌های سه بعدی و حتی در طراحی پروتئین استفاده شده‌اند. این نتایج تجربی قوی پتانسیل مدل‌های مولد مبتنی بر امتیاز را برجسته می‌کند. با این حال، آنها در حال حاضر **فاقد ضمانت‌های آماری دقیق در مورد توزیع نمونه‌های تولید شــده هســتند،** که مانع اســتقرار ایمن آنها در ســناریوهای پر خطر می‌شود. به عنوان مثال، یک رادیولوژیســت را در نظر بگیرید که یک توموگرافی کامپیوتری (CT) اسکن شکم بیمار را که از طریق یک مدل تولیدی مبتنی بر امتیاز بازسازی شده است، در نظر بگیرید. چقدر باید از جزئیات دقیق تصویر ارائه شده اطمینان داشته باشند. به عبارت دیگر، نمونه‌های آینده چقدر با تصویر ارائه شده متفاوت خواهند بود و چقدر می‌توانیم انتظار داشته باشیم که آنها از تصویر حقیقت زمینه و داده‌ها فاصله داشته باشند؟ در این کار ما بر روی مشکلات رگرسیون تصویر به تصویر تمرکز می‌کنیم، جایی که ما علاقه مند به بازیابی یک تصویر حقیقت زمین با کیفیت بالا با توجه به یک مشاهده با کیفیت پایین هستیم.

[7] Reverse-time stochastic differential equation

۲-۲- مفاهیم ارائه‌شده در مقاله

۲-۲-۱- معادله دیفرانسیل تصادفی

عبارتی که معادله دیفرانسـیل تصـادفی[۸] (SDE) را برای نمونه‌گیری شـرطی مبتنی بر امتیاز شـامل اصـطلاحات رانش[۹] و انتشار[۱۰] نشان می‌دهد بدین شکل است:

$$d\boldsymbol{x} = [h(\boldsymbol{x},t) - g(t)^2 \nabla_x log p_t(\boldsymbol{x})]dt + g(t)d\overline{w}$$

در این عبارت:

x نشان دهنده نمونه در حال به روز رسانی است.

$h(\boldsymbol{x},t)$ عبارت رانش را نشان می‌دهد و نمونه را به سمت اطلاعات شرطی هدایت می‌کند.

$g(t)$ یانگر عبارت انتشار است که پخش نویز را در طول فرآیند نمونه برداری کنترل می‌کند.

$\nabla_x log p_t(\boldsymbol{x})$تابع امتیاز را نشـان می‌دهد و اطلاعاتی در مورد گرادیان احتمال ورود به سیسـتم با توجه به نمونه ارائه می‌کند.

$d\overline{w}$یک فرآیند وینر را نشـان می‌دهد، یک اصـطلاح نوسـان تصـادفی، که اغلب در معادلات دیفرانسـیل تصـادفی برای مدل‌سازی نویز یا نوسانات تصادفی استفاده می‌شود.

این عبارت پویایی نمونه‌گیری شـرطی مبتنی بر امتیاز را نشـان می‌دهد، جایی که نمونه xدر طول زمان مطابق با عبارت رانش $h(\boldsymbol{x},t)$ به‌روزرسانی می‌شـود. گرادیان لگ احتمال، و عبارت انتشـار $g(t)$ نیز نشـان دهنده ماهیت تصادفی فرآیند نمونه برداری است.

می‌توان با دانستن p(y|x) روش نمونه گیری را بر اساس مشاهده y شرطی کرد. بدین صورت:

$$\nabla_x \, log \, p_t(x|y) \; = \nabla_x \, log \, p_t(y|x) + \nabla_x \, log \, p_t(x)$$

[8] Stochastic Differential Equation

[9] Drift

[10] Diffusion

پیشرفت‌های اخیر در مدل‌سازی مولد توسط سانگ و ارمون نشان داد که می‌توان یک شبکه امتیاز شرطی زمانی $s(\tilde{x}, t)$ را برای تقریب نمره $\nabla_x \log p_t(\tilde{x})$ از طریق تطبیق امتیاز زدایی به طور موثر آموزش داد. به این ترتیب، با توجه به یک SDE زمان جلو که فرآیند مشاهده را مدل می‌کند، یک شبکه امتیاز $s(\tilde{x}, t) \approx \nabla_x \log p_t(\tilde{x})$ و عبارت احتمال $p(y|\tilde{x})$، می‌توان نمونه از $p(x|y)$ با حل SDE زمان معکوس شرطی با هر گسسته سازی (به عنوان مثال، اویلر-مارویاما) یا طرح پیش بینی کننده- تصحیح کننده تعریف کرد.

۲-۲-۲- پیش‌بینی منسجم[11]

همانطور که در بخش ۳-۱ صحبت شد، باید پیش‌بینی داشته باشیم. پیش بینی منسجم شامل روش‌شناسی‌های مختلفی برای ایجاد ضمانت‌های عدم قطعیت با نمونه محدود و آماری معتبر برای پیش‌بینی‌کننده‌های عمومی‌بدون ایجاد هرگونه فرضی در مورد توزیع پاسخ است (یعنی آن‌ها بدون توزیع هستند).

شفر و ووک در کار خود چارچوبی به نام احتمال نظری بازی[12] (GTP) پیشنهاد می‌کنند که تعمیم نظریه احتمال کلاسیک است. در این چارچوب، آنها مفهوم پوشش[13] را برای کمی‌کردن سطح باور یا اطمینان در پیش‌بینی انجام شده با روش آماری معرفی می‌کنند. پوشش نشان دهنده نسبت رویدادها یا نتایج آینده است که در محدوده یا بازه پیش بینی شده قرار می‌گیرند. به عبارت دیگر، اگر یک روش پیش‌بینی ادعا می‌کند که یک بازه پیش‌بینی با پوشش ۹۵ درصدی ارائه می‌کند، به این معنی است که در درازمدت، ۹۵ درصد از نتایج واقعی باید در آن بازه قرار گیرند.

$$\mathbb{P}[z_{m+1} \in \mathcal{C}] \geq 1 - \alpha.$$

مفهوم پوشش به ویژه در پیش‌بینی و پیش‌بینی اهمیت دارد، زیرا معیاری از قابلیت اطمینان یا دقت پیش‌بینی‌ها را ارائه می‌دهد. یک روش پیش‌بینی با پوشش بالا قابل اعتمادتر در نظر گرفته می‌شود، زیرا به طور مداوم بخش بزرگی از نتایج واقعی را در محدوده پیش‌بینی‌شده خود ثبت می‌کند.

[11] Conformal Prediction

[12] Game-theoretic Probability

[13] Coverage

در بسیاری از سناریوها (به عنوان مثال، رگرسیون)، مفهوم طبیعی عدم قطعیت ممکن است با پوشش نادرست همانطور که در بالا توضیح داده شد متفاوت باشد (به عنوان مثال، $\ell 2$).

۲-۲-۳- کنترل ریسک منسجم[۱۴]

کنترل ریسک منسجم مفهومی‌است که از حوزه پیش بینی منسجم نشات می‌گیرد، چارچوبی در یادگیری ماشین و استنتاج آماری که راهی برای کمی‌سازی عدم قطعیت و کنترل ریسک در مدل‌های پیش بینی ارائه می‌دهد. در یادگیری ماشینی سنتی، مدل‌ها معمولا پیش‌بینی‌های نقطه‌ای را بدون ارائه هیچ معیار عدم قطعیت ارائه می‌کنند. با این حال، در بسیاری از برنامه‌های کاربردی دنیای واقعی، نه تنها پیش‌بینی دقیق، بلکه تعیین کمیت عدم قطعیت مرتبط با آن پیش‌بینی‌ها نیز مهم است.

پیش‌بینی منسجم این نیاز را با ارائه مناطق پیش‌بینی (مجموعه‌ها) به جای پیش‌بینی نقطه‌ای برطرف می‌کند. این مناطق پیش‌بینی با تضمین‌های قابل اثبات اعتبار ارائه می‌شوند، به این معنی که نتیجه واقعی با احتمال خاصی در منطقه پیش‌بینی قرار می‌گیرد. پیش‌بینی مطابق روشی اصولی برای کنترل ریسک مرتبط با پیش‌بینی‌ها فراهم می‌کند. بنابراین، کنترل ریسک منسجم به فرآیند مدیریت و کنترل ریسک مرتبط با مدل‌های پیش‌بینی در چارچوب پیش‌بینی منسجم اشاره دارد. این مورد شامل اطمینان از این است که مناطق پیش‌بینی تولید شده توسط مدل دارای ویژگی‌های مورد نظر از نظر پوشش و دقت هستند.

ایده کلیدی در پس کنترل ریسک منسجم، ارائه تضمین‌هایی است که نتیجه واقعی در مناطق پیش بینی با احتمال بالا قرار می‌گیرد، و به کاربران اجازه می‌دهد تا تصمیمات آگاهانه بگیرند و در عین حال عدم اطمینان را در نظر بگیرند. این امر به ویژه در کاربردهایی که عواقب پیش‌بینی‌های نادرست می‌تواند شدید باشد، مانند تشخیص پزشکی، امور مالی و سیستم‌های حیاتی از اهمیت ویژه‌ای برخوردار است.

[14] Conformal risk control

فرض کنید که $\mathcal{S}_{\text{cal}} = \{(x_i, y_i)\}_{i=1}^{n} \sim \mathcal{D}^n$ یک مجموعه کالیبراسیون از n i.i.d باشد. (نمونه‌هایی از یک توزیع ناشناخته D بر X × Y). برای سطح ریسک مطلوب ε ≥ 0 و احتمال شکست $\delta \in [۰,۱]$ یک $\mathcal{I}: \mathcal{Y} \to \mathcal{X}' \subseteq 2^{\mathcal{X}}$ یک تابع خطای RCPS از نوع w.r.t پیش‌بینی‌کننده تصادفی با مقدار مجموعه است.

$$\ell: \mathcal{X} \times \mathcal{X}' \to \mathbb{R}$$

$$\mathbb{P}_{\mathcal{S}_{\text{cal}}}\left[\mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(x, \mathcal{I}(y))] \le \epsilon\right] \ge 1 - \delta.$$

بیتس و همکاران اولین روش کنترل ریسک منسجم را برای توابع خطای یکنواخت غیرافزاینده معرفی کرد، آنهایی که برای x ثابت، خواهیم داشت:

$$\mathcal{I}(y) \subset \mathcal{I}'(y) \Rightarrow \ell\left(x, \mathcal{I}'(y)\right) \le \ell\left(x, \mathcal{I}(y)\right).$$

به این ترتیب افزایش لندازه مجموعه‌ها نمی‌تولند ارزش خطا را افزایش دهد. علاوه بر این، فرض کنید برای یک ورودی ثابت y، خانواده پیش‌بینی‌کننده‌های مجموعه $\{\mathcal{I}_\lambda(y)\}_{\lambda\in\Lambda}$، نمایه‌سازی شده توسط $\lambda \in \Lambda, \Lambda \subset \bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$، ویژگی تودرتو زیر را برآورده می‌کند:

$$\lambda_1 < \lambda_2 \Rightarrow \mathcal{I}_{\lambda_1}(y) \subset \mathcal{I}_{\lambda_2}(y).$$

$R(\lambda) = \mathbb{E}[\ell(x, \mathcal{I}_\lambda(y))]$ ریسک $\mathcal{I}_\lambda(y)$ و $\widehat{R}(\lambda)$ تخمین تجربی آن را در یک مجموعه کالیبراسیون مشخص کنید. $\mathcal{S}_{\text{cal}} = \{(x_i, y_i)\}_{i=1}^{n}$. در نهایت، اجازه دهید $\widehat{R}^{+}(\lambda)$ یک کران اطمینان بالای نقطه‌ای[15] (UCB) باشد که ریسک را پوشش می‌دهد، یعنی

$$\mathbb{P}\left[R(\lambda) \le \hat{R}^{+}(\lambda)\right] \ge 1 - \delta$$

برای هر یک، مقدار ثابت λ - که می‌تواند با استفاده از نابرابری‌های تمرکز به دست آید، نشان می‌دهد که انتخاب

$$\hat{\lambda} = inf\left\{\lambda \in \Lambda: \widehat{R}^{+}\left(\lambda'\right) < \epsilon, \forall\lambda' \ge \lambda\right\}$$

[15] Upper Confidence Bound

تضـــمین می‌کند که $\mathcal{I}_{\lambda}(y)$ یک RCPS- (ε, δ) اســـت. به عبارت دیگر، انتخاب $\hat{\lambda}$ به عنوان کوچکترین λ به طوری که UCB زیر سطح مورد نظر ε برای همه مقادیر $\lambda \geq \hat{\lambda}$ باشد، سطح ریسک ε را با احتمال حداقل ۱-δ کنترل می‌کند. (یک مقدار متوجه شدن این مفهوم سخت است).

۳- الگوریتم مقاله

به‌جای یک پیش‌بینی‌کننده نقطه $f: Y \to X$، فرض می‌کنیم که به روش نمونه‌گیری تصـــادفی $F: Y \to X$ دسترسی داریم به طوری که F(y) یک متغیر تصادفی با توزیع ناشناخته Qy است که امیدواریم توزیع پسین x را با y تقریب کند، یعنی $Qy \approx p(x|y)$. با این حال، ما هیچ فرضی در مورد کیفیت این تقریب برای حفظ نتایج خود نداریم. همانطور که توضیح داده شد، F را می‌توان با استفاده از یک شبکه امتیاز شرطی زمانی $s(\tilde{x}, t)$ و یک SDE زمان معکوس به دست آورد. می‌توان سه منبع مجزای تصادفی را در یک مسئله رگرسیون تصویر به تصویر کلی تصادفی شناسایی کرد:

۱. p(x) ناشناخته قبلی در فضای تصاویر حقیقت زمین، به صورت $x \sim p(x)$،

۲. تصادفی بودن در فرآیند مشاهده y (که می‌تواند توسط یک SDE زمان جلو بر روی x مدل شود)،

۳. ابتدا تضـمین‌های پیش‌بینی منسـجم را برای یک مشـاهده ثابت y ارائه می‌کنیم، و سـپس به سـمت کنترل ریسک منسجم برای تصویر حقیقی x می‌رویم.

۳-۱- پوشش ورودی[۱۶]

پوشـــش ورودی به ویژگی بازه‌ها یا مجموعه‌های پیش‌بینی در اســـتنتاج آماری اشـــاره دارد که در آن هر بازه یا مجموعه پیش‌بینی فردی تضمین‌های پوشش قابل اعتمادی را برای پیش‌بینی مربوطه خود ارائه می‌کند.

[16] Entrywise coverage

در زمینه پیش‌بینی منســـجم یا رگرســـیون کمی، پوشـــش ورودی به این معنی اســـت که برای هر پیش‌بینی انجام‌شـــده، بازه یا مجموعه پیش‌بینی مرتبط احتمال زیادی دارد که حاوی نتیجه واقعی باشـــد. به عبارت دیگر، پوشـــش ورودی تضـــمین می‌کند که فواصـــل پیش‌بینی برای هر پیش‌بینی منفرد معتبر اســـت، نه فقط به طور میانگین یا در مجموع.

از نظر ریاضی، اگر مجموعه ای از n پیش بینی با فواصل پیش بینی متناظر S_i و نتایج واقعی y_i داشـته باشـیم، پوشش ورودی به این معنی است که برای هر i داریم:

$$P(y_i \in S_i) \geq 1 - \alpha$$

که در آن α سـطح معنی داری اسـت که نشـان دهنده احتمال قرار گرفتن نتیجه واقعی در خارج از فاصـله پیش بینی است.

۳-۲- پوشش ورودی تضمینی چندک‌های کالیبره شده[17]

این روش نتیجه ای در چارچوب پیش بینی منســـجم اســـت، روشـــی در یادگیری ماشـــین و اســـتنتاج آماری که مجموعه‌های پیش بینی معتبر با نرخ خطای شـــناخته شـــده را ارائه می‌دهد. این لم به طور خاص به مفهوم کالیبراســـیون در زمینه فواصـــل پیش‌بینی مبتنی بر چندت می‌پردازد. در پیش‌بینی منســـجم، کالیبراســـیون به خاصیتی اشاره دارد که سطح پوشش پیش‌بینی‌شده (نسبت نتایج واقعی تحت پوشش فواصل پیش‌بینی) با سطح اطمینان مشـــخص شـــده مطابقت دارد. به عبارت دیگر، زمانی گفته می‌شـــود که یک فاصـــله پیش‌بینی کالیبره می‌شـــود که برای یک ســـطح اطمینان مشـــخص (مثلاً ۹۵٪)، نســـبت دفعاتی که نتیجه واقعی در این بازه قرار می‌گیرد به سطح اطمینان نزدیک باشد.

[17] Calibrated Quantiles Guarantee Entrywise Coverage

این روش بیان می‌کند که اگر مدل رگرسیون چندکی که برای ساخت فواصل پیش‌بینی استفاده می‌شود، کالیبره شده باشد، فواصل پیش‌بینی حاصل تضمین‌های پوشش ورودی را ارائه می‌کنند. پوشش ورودی به این معنی است که برای هر بازه پیش‌بینی فردی، نتیجه واقعی در بازه زمانی با حداقل سطح اطمینان مشخص قرار می‌گیرد. پوشش ورودی اینگونه تعریف می‌شود:

$$\mathcal{I}_{\lambda}^{\alpha}(y) = \left[\hat{l}_{j,\alpha}, \hat{u}_{j,\alpha}\right]$$

فواصل در $\mathcal{I}_{\lambda}^{\alpha}(y)$ وابسته به ویژگی هستند و مناطقی از تصویر را می‌گیرند که فرآیند نمونه برداری F(y) ممکن است عدم قطعیت بیشتری داشته باشد. فواصل در $\mathcal{I}_{\lambda}^{\alpha}(y)$ از نظر آماری برای هر تعداد نمونه m و هر توزیع Qy معتبر هستند، یعنی یک مفهوم اکتشافی از عدم قطعیت نیستند که می‌تواند فرآیندی زمان بر و ظریف باشد، به خصوص زمانی که اندازه تصاویر قابل توجه باشد. از سوی دیگر، ساخت فواصل $\mathcal{I}_{\lambda}^{\alpha}(y)$مستلزم نمونه برداری به تعداد کافی از F(y) است که ممکن است سخت باشد اما توجه داشته باشید که نمونه‌برداری کاملا موازی‌پذیر است و بنابراین اگر تعداد بیشتری گره محاسباتی در دسترس باشد، هیچ پیچیدگی اضافی ایجاد نمی‌شود.

۳-۳- تعمیم کنترل ریسک منسجم: K-RCPS

به طور طبیعی، فرد می‌خواهد یک روش کنترل ریسک منسجم مناسب برای کوتاه‌ترین فاصله زمانی ممکن باشد. ابتدا باید شدت پیکسل‌ها را بین [۰، ۱] نرمال کنید و تابع خطا را اینگونه در نظر بگیرید.

$$\ell^{01}\big(x, \mathcal{I}(y)\big) = \frac{1}{d}\sum_{j\in[d]} \mathbb{1}\left[x_j \notin \mathcal{I}(y)_j\right],$$

این تابع تعداد (متوسط) پیکسل‌های حقیقی را که خارج از فواصل مربوطه در I(y) قرار می‌گیرند، می‌شمارد. پیش‌بینی‌کننده با ارزش مجموعه ثابت $\mathcal{U}(y) = [0,1]^d$ ریسک را کنترل می‌کند، یعنی $R^{01}(\lambda) = \mathbb{E}\left[\ell^{01}(x, \mathcal{U}(y))\right] = 0$. اما، چنین پیش‌بینی‌کننده‌ای خوب نیست. در نظر بگیرید $\{\mathcal{I}_{\lambda}(y)\}_{\lambda\in\Lambda}, \Lambda \subset$ -
$\mathbb{R}$خانواده ای از پیش بینی کننده‌ها باشد که ویژگی تودرتو در RCPS را برآورده می‌کند. به طور خاص، مقاله پارامترسازی افزایشی زیر را در λ پیشنهاد می‌کنیم.

$$\mathcal{I}_\lambda(y)_j = \left[\hat{l}_j - \lambda, \hat{u}_j + \lambda\right]$$

برای برخی از نقاط پایانی پایین و بالایی $\hat{l}_j < \hat{u}_j$ که ممکن است به y بستگی داشته باشند. برای این خانواده خاص انتخاب شده از پیش بینی‌های تودرتو، نتیجه می‌شود که میانگین طول بازه است:

$$\bar{I}(\lambda) = \frac{1}{d}\sum_{j\in[d]} \left(\hat{u}_j - \hat{l}_j\right) + 2\lambda,$$

علاوه بر این، می‌توانیم $\hat{l}_j$ و $\hat{u}_j$را به عنوان چندک‌های کالیبره‌شده با پوشش ورودی، مثال کنیم، یعنی $\mathcal{I}_\lambda^\alpha(y) = \left[\hat{l}_{j,\alpha} - \lambda, \hat{u}_{j,\alpha} + \lambda\right]$ برای چنین دسته‌ای از پیش‌بینی‌کننده‌ها - از آنجایی که تلفات $\ell 01$ به طور یکنواخت غیرافزاینده است - روش اصلی RCPS معادل مسئله بهینه‌سازی محدود زیر است.

$$\hat{\lambda} = \arg\min_{\lambda\in\Lambda} \bar{I}(\lambda)$$

$$\text{s.t. } \hat{R}^{01+}(\lambda') < \epsilon, \forall\lambda' \geq \lambda (\mathrm{P}_1)$$

که به طور طبیعی λ را به حداقل می‌رساند. با این حال، همانطور که در شکل ۱ نشان داده شده است، بهینه سازی طول متوسط بازه روی یک پارامتر اسکالر λ به طور کلی کمتر از حد بهینه است. در این شکل مقیاس‌ها اینگونه اند:

$$.(\mathcal{I}_{\boldsymbol{\lambda}})_j = [-1 - \lambda_j, 1 + \lambda_j], \boldsymbol{\lambda} = (\lambda_1, \lambda_2) \text{ و } \mathcal{S}_{\text{cal}} \sim \mathcal{N}(\mu, \mathbb{I}_2)^n, n = 128$$

برای $\hat{R}^{01+}(\boldsymbol{\lambda})$ $\epsilon = \delta = 0.1,$از طریق هیبریداسیون Hoeffding-Bentkus به دست می‌آید. مناطق سبز مناطقی را نشان می‌دهند که $\hat{R}^{01+}(\boldsymbol{\lambda}) \leq \epsilon$ ، و برعکس برای مناطق قرمز. (a) نشان می‌دهد که وقتی ویژگی‌ها به طور متقارن در اطراف فواصل متمرکز می‌شوند، کمینه کردن λ1 = λ2 = λ (ستاره آبی) میانگین طول بازه را به حداقل می‌رساند، در حالی که (b) نشان می‌دهد که در حالت کلی، λ بهینه (ستاره نارنجی) ممکن است λ1نابرابر با λ2 داشته باشد. Δ افزایش میانگین فاصله بازه به دست آمده با انتخاب ستاره نارنجی به جای ستاره آبی را برجسته می‌کند.

$$\mathcal{I}_{\boldsymbol{\lambda}}(y)_j = \left[\hat{l}_j - \lambda_j, \hat{u}_j + \lambda_j\right],$$

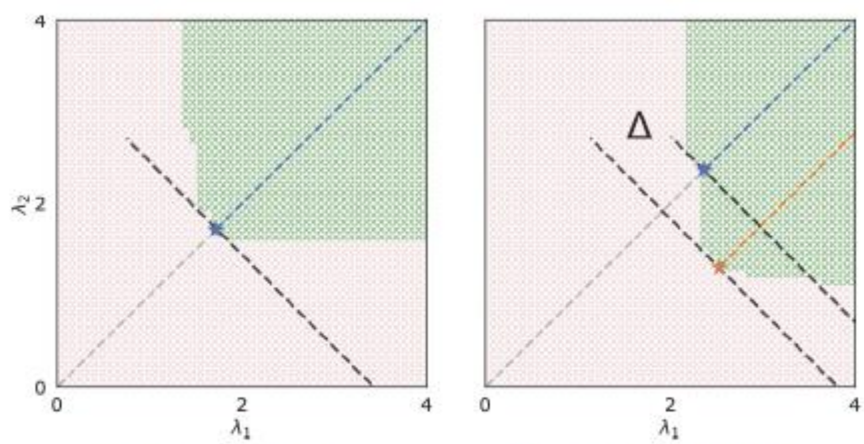
که از آن می‌توان تابع معادل $\bar{I}(\boldsymbol{\lambda})$ را تعریف کرد. به طور خاص، با اســتفاده از فواصــل کالیبره شــده مانند قبل، تعریف کنید

$$\mathcal{I}_{\lambda}^{\alpha}(y) = \left[\hat{l}_{j,\alpha} - \lambda_j, \hat{u}_{j,\alpha} + \lambda_j\right]$$

اکنون توجه داشــته باشــید که $\ell^{01}(x, \mathcal{I}_{\boldsymbol{\lambda}}(y))$ به طور یکنواخت غیرافزاینده است. بنابراین، برای یک بردار ثابت $\eta \in \mathbb{R}^d$در $\eta \geq 0$، در ورودی، رابطه را گسترش می‌دهد.

$$\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}\in\Lambda^d} \sum_{j\in[d]} \lambda_j \text{ s.t. } \widehat{R}^{01+}(\boldsymbol{\lambda} + \beta\boldsymbol{\eta}) < \epsilon, (\mathrm{P}_d), \forall\beta \geq 0$$

بطور شــهودی، $\hat{\boldsymbol{\lambda}}$ مجموع ورودی‌های خود را به حداقل می‌رســاند به طوری که UCB برای همه نقاط ســمت راست آن در امتداد جهت η که توسط β پارامتر شده است، کوچکتر از ε باشد ما اکنون یک نتیجه کلی کنترل ریسـک با ابعاد بالا را نشـان می‌دهیم که برای هر تابع زیان یکنواخت بدون افزایش ℓ (و نه فقط ℓ^{01}همانطور که در رابطه بالا می‌بینید) با ریسک R(λ)، تخمین تجربی $\widehat{R}(\boldsymbol{\lambda})$ و $\widehat{R}^{+}(\boldsymbol{\lambda})$ UCBمربوطه صادق است.

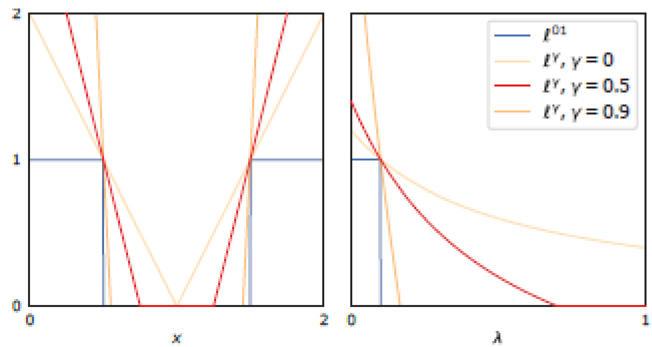


(a) $\mu = [-1, 1]^T$. (b) $\mu = [-2, 0.75]^T$.

شکل ۳: نمایش تصویری از عدم بهینه بودن انتخاب یک پارامتر اسکالر منفرد λ w.r.t میانگین طول بازه.

مسئله بهینه سازی مربوط به قضیه کنترل ریسک طول متوسط بازه بهینه را می‌توان به صورت زیر فرموله کرد:

$$\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}\in\Lambda^d} \sum_{j\in[d]} \lambda_j \text{ s.t. } \widehat{R}^{+}(\boldsymbol{\lambda} + \beta\boldsymbol{\eta}) < \epsilon$$



(a) ℓ^{01}, ℓ^{γ} as a function of x. (b) ℓ^{01}, ℓ^{γ} as a function of λ.

شکل ۴: نمایشی از $\ell^{01}(x, \mathcal{J}_\lambda(y))$ و $\ell^{\gamma}(x, \mathcal{J}_\lambda(y))$ الف: $\lambda = 0$ و ب) $x = 1.6$

همانطور که در شکل ۴ که نمایشی از $\ell^{01}(x, \mathcal{J}_\lambda(y))$ و $\ell^{\gamma}(x, \mathcal{J}_\lambda(y))$ در $\mathcal{J}_\lambda(y) = [0.50 - \lambda, 1.50 + \lambda], \gamma \in \{۰,۰.۵,۰.۹\}$ مشاهده می‌نمایید، ℓ^{01} در λ محدب نیست. از این رو، حل بهینه به طور مستقیم امکان پذیر نیست و باید آن را به یک مسئله بهینه‌سازی محدب با استفاده از یک کران بالایی محدب تبدیل کنیم.

$$\ell^{\gamma}\left(x, \mathcal{J}_{\boldsymbol{\lambda}}(y)\right) = \frac{1}{d} \sum_{j \in [d]} \left[\frac{2(1+q)}{I(\boldsymbol{\lambda})_j} \left|x_j - c_j\right| - q\right]_+ ,$$

$$q = \gamma/(1-\gamma), \gamma \in [0,1), I(\boldsymbol{\lambda})_j = \hat{u}_j - \hat{l}_j + ۲\lambda_j, c_j = \left(\hat{u}_j + \hat{l}_j\right)/2, [\cdot]_+ = max(0,\cdot)$$

در انتها مقاله این تابع را برای محاسبه $\hat{\boldsymbol{\lambda}}$ معرفی می‌نماید.

$$\tilde{\boldsymbol{\lambda}}_K = \underset{\boldsymbol{\lambda} \in \Lambda^K}{\arg min} \sum_{k \in [K]} n_k \lambda_k \text{ s.t. } \widehat{R}^{\gamma}(M\boldsymbol{\lambda}) \leq \epsilon, (\mathrm{P}_K)$$

برای هر پارتیشن K تعریف شده توسط کاربر از ویژگی‌های [d] که می‌توانند با ماتریس عضویت $M \in \{0,1\}^{d \times K}$ شناسایی شوند که در آن هر ویژگی به (فقط) یکی از گروه‌های K با $n_k := \left|\{j \in [d]: M_{jk} = 1\}\right|, \sum_{k \in [K]} n_k = d$ تعلق دارد.

این تابع را با استفاده از ابزارهای بهینه‌سازی مانند CVXPY برای مسئله‌های مختلف حل نمود. الگوریتم K-RCPS را در الگوریتم ۱ مشاهده می‌نمایید.

Algorithm 1 K-RCPS

Input: risk level $\epsilon \geq 0$, failure probability $\delta \in [0,1]$, calibration set $\mathcal{S}_{\text{cal}} = \{(x_i, y_i)\}_{i=1}^{n}$ of n i.i.d. samples such that $n = n_{\text{opt}} + n_{\text{RCPS}}$, membership function $\mathcal{M}$, family of set-valued predictors $\{\mathcal{I}_{\boldsymbol{\lambda}}(y) = [\hat{l}_j - \lambda_j, u_j + \lambda_j]\}_{\boldsymbol{\lambda} \in \Lambda^d}$, initial (large) value $\beta_{\max}$, stepsize $\mathrm{d}\beta > 0$.

Split $\mathcal{S}_{\text{cal}}$ into $\mathcal{S}_{\text{opt}}, \mathcal{S}_{\text{RCPS}}$

$M \leftarrow \mathcal{M}(\mathcal{S}_{\text{opt}})$

$\widetilde{\boldsymbol{\lambda}}_K \leftarrow \texttt{SOLVE-PK}(\mathcal{S}_{\text{opt}}, M)$

$\boldsymbol{\lambda} \leftarrow M\widetilde{\boldsymbol{\lambda}}_K + \beta_{\max}\mathbf{1}$

$\hat{R}^{01+}(\boldsymbol{\lambda}) \leftarrow 0$

while $\hat{R}^{01+}(\boldsymbol{\lambda}) \leq \epsilon$ **do**

  $\boldsymbol{\lambda}_{\text{prev}} \leftarrow \boldsymbol{\lambda}$

  $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} - (\mathrm{d}\beta)\mathbf{1}$

  $\boldsymbol{\lambda} \leftarrow [\boldsymbol{\lambda}]_+$

  $\hat{R}^{01}(\boldsymbol{\lambda}) \leftarrow 1/n_{\text{RCPS}} \cdot \sum_{(x_i,y_i) \in \mathcal{S}_{\text{RCPS}}} \ell^{01}(x_i, \mathcal{I}_{\boldsymbol{\lambda}}(y_i))$

  $\hat{R}^{01+}(\boldsymbol{\lambda}) \leftarrow \text{UCB}(n_{\text{RCPS}}, \delta, \hat{R}^{01}(\boldsymbol{\lambda}))$

end while

$\hat{\boldsymbol{\lambda}}_K \leftarrow \boldsymbol{\lambda}_{\text{prev}}$

return $\hat{\boldsymbol{\lambda}}_K$

۴- آزمایش این الگوریتم

مقاله به نشان دادن مزیت K-RCPS از نظر طول متوسط فاصله در دو مشکل حذف نویز با ابعاد بالا در دنیای واقعی پرداخته است: یکی در تصاویر طبیعی از چهره‌ها و همچنین در سی تی اسکن شکم. K-RCPS را با الگوریتم اصلی RCPS در چندین مفهوم پایه عدم قطعیت برای مسئله رفع نویز کردن روی دو مجموعه داده celebA و AbdomenCT-1K مقایسه می‌کنیم:

- quantile regression
- MC-Dropout

- N-Conffusion
- naive

که این مفاهیم را چون خارج از بحث اصلی ما هستند، از توضیحشان صرف نظر می‌کنیم و تنها قصد داریم تا نشان دهیم که الگوریتم پیشنهادی پاسخ خوبی در این مفاهیم در دادگان اشاره شده داشته است. شکل ۵ نمایشی از مقایسه همه مفاهیم عدم قطعیت با RCPS و K RCPS از نظر تضمین‌های ارائه شده می‌باشد. برای هر دو مجموعه داده، K-RCPS کمترین فواصل را در بین روش‌هایی فراهم می‌کند که هم پوشش ورودی و هم کنترل ریسک را برای مدل‌های انتشار فراهم می‌کند. هنگامی‌که محدودیت‌های پوشش ورودی را کاهش می‌دهد، K-RCPS همچنان کمترین فواصل را ارائه می‌دهد. در میان روش‌های کمی‌سازی عدم قطعیت که به یک مدل انتشار مربوط نمی‌شوند، K-RCPS با نمونه‌برداری ساده نتایج بهتری را در مجموعه داده CelebA ارائه می‌کند.

Uncertainty	Diffusion model?	Entrywise coverage?	Risk control?	Calibration procedure	Mean interval length CelebA	Mean interval length AbdomenCT-1K
QR	✗	✗	✓	RCPS	0.4843 ± 0.0121	0.2943 ± 0.0060
MC-Dropout	✗	✗	✓	RCPS	0.6314 ± 0.0109	0.2810 ± 0.0013
N-Con$ffusion$						
— multiplicative	✗	✗	✓	RCPS	0.6949 ± 0.0084	0.1126 ± 0.0020
— additive	✗	✗	✓	RCPS	0.3314 ± 0.0040	0.1164 ± 0.0024
— additive	✗	✗	✓	K-RCPS	0.3131 ± 0.0056	0.1136 ± 0.0019
Naive quantiles	✓	✗	✓	RCPS	0.2688 ± 0.0068	0.1518 ± 0.0016
Naive quantiles	✓	✗	✓	K-RCPS	0.2523 ± 0.0052	0.1374 ± 0.0019
Calibrated quantiles	✓	✓	✓	RCPS	0.2762 ± 0.0059	0.1506 ± 0.0014
Calibrated quantiles	✓	✓	✓	K-RCPS	0.2644 ± 0.0067	0.1369 ± 0.0016

شکل ۵: مقایسه همه مفاهیم عدم قطعیت با RCPS و K RCPS از نظر تضمین‌های ارائه شده و میانگین طول بازه بیش از ۲۰ ترسیم مستقل از Scal.

۵- خلاصه کار

مراحل انجام شده در مقالهبه صورت زیر می‌باشد.

۱. با توجه به یک شبکه امتیاز ثابت، یک مشاهده با کیفیت پایین و هر روش نمونه‌گیری، نشان داده شد که چگونه فواصل کالیبره‌شده ورودی معتبری ایجاد می‌شوند که پوشش نمونه‌های آینده را فراهم می‌کند، یعنی نمونه‌های آینده در بازه‌های با احتمال زیاد قرار می‌گیرند.

۲. یک روش جدید کنترل ریسک منسجم با ابعاد بالا معرفی شد که میانگین طول بازه را مستقیما به حداقل می‌رساند. در حالی که تضمین تعداد پیکسل‌هایی که در تصویر حقیقتی خارج از این فواصل قرار می‌گیرند، در مشاهدات با کیفیت پایین و غیرقابل مشاهده آتی کمتر از سطح مشخص شده توسط کاربر است.

روش جدید کنترل ریسک منسجم ارائه‌شده، K-RCPS، بردار $\tilde{\boldsymbol{\lambda}}_K$ را پیدا می‌کند که یک راه حل برای مسئله بهینه‌سازی غیر محدب زیر

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}\in\Lambda^d}{\arg min} \sum_{j\in[d]} \lambda_j \text{ s.t. } \widehat{R}^{+}(\boldsymbol{\lambda} + \beta\boldsymbol{\eta}) < \epsilon$$

را از طریق یک روش دو مرحله ای تقریب می‌زند:

۱) ابتدا راه حل بهینه $\tilde{\boldsymbol{\lambda}}_K$ را برای یک مسئله تعریف شده بدست آورید(با استفاده از ابزارهای بهینه‌سازی)

$$\tilde{\boldsymbol{\lambda}}_K = \underset{\boldsymbol{\lambda}\in\Lambda^K}{\arg min} \sum_{k\in[K]} n_k\lambda_k \text{ s.t. } \widehat{R}^{\gamma}(M\boldsymbol{\lambda}) \leq \epsilon, (\mathrm{P}_K)$$

۲) انتخاب $\hat{\beta} \in \Lambda$ به گونه ای که

$$\hat{\beta} = inf\left\{\beta \in \Lambda : \widehat{R}^{01+}\left(M\tilde{\lambda}_K + \beta' 1\right) < \epsilon, \forall\beta' \geq \beta\right\}$$
$$\text{and return } \hat{\lambda}_K = M\tilde{\lambda}_K + \hat{\beta}1.$$

به طور مستقیم، الگوریتم K-RCPS معادل اجرای روش اصلی RCPS در امتداد خط $M\tilde{\lambda}_K + \beta 1$است که توسط β پارامتر شده است.

۳. رویکرد برای حذف نویز از تصاویر طبیعی صورت و همچنین برای توموگرافی کامپیوتری شکم مقایسه شد و به نتایج پیشرفته‌ای در طول متوسط فاصله دست می‌یابیم. نتایج نشان دادند که این فواصل دقیقاً نوع بافتی را که می‌توان توسط مدل بازسازی کرد مشخص می‌کند.