400104964

آیلین رسته

(1) الف) در یادگیری آفلاین به دلیل distribution shift ممکن است برخی state-action های مشاهده نشوند و مقدار تخمین زده شده برای دیگر آنها توسط مدل بسیار بیشتر از مقدار واقعی آنها شود و این خطا propagate شود. به این منظور با عبارت $E_{\mu}[Q]$ را بیشینه می‌کنیم تا مقادیر Q ای که بزرگ هستند را کاهش دهیم و از این شکل جلوگیری کنیم اما این خودش یک مشکل جدید اضافه می‌کند که با این کار ویژگی‌هایی که واقعا بزرگ هستند و در دیتاست موجود بودند هم کاهش می‌یابد. برای همین عبارت $E_{\hat{\pi}_\beta}[Q]$ را از آن کم می‌کنیم تا مانع از کم شدن ویژگی‌هایی شویم که به موردی مورد بررسی در بخش distribution shift ندارند.

ب) ابتدا بهینه‌سازی داخلی زیر را حل می‌کنیم.

$$\max_{\mu} E_{s\sim D}\left[E_{a\sim\mu(a|s)}[\alpha Q(s,a)] - H(\mu(.|s))\right]$$

$$\text{s.t.} \quad \forall s: \sum_a \mu(a|s) = 1$$

می‌توانیم برای هر s مسئله را جداگانه حل کنیم پس مسئله بهینه‌سازی به فرم زیر تبدیل می‌شود:

$$\max_{\mu} E_{a\sim\mu(s,a)}[\alpha Q(s,a)] + H(\mu(.|s))$$

$$\text{s.t.} \quad \sum_a \mu(a|s) = 1$$

مسأله را به کمک لاگرانژین آن حل می‌کنیم.

$$L(p,\lambda) = \sum_a p(a|s)\big(Q(s,a) - \log p(a|s)\big) + \lambda\Big(\sum_a p(a|s) - 1\Big)$$

$$\frac{\partial L(p,\lambda)}{\partial p(a|s)} = \alpha Q(s,a) - \log p(a|s) - 1 + \lambda = 0$$

$$\Rightarrow p(a|s) = e^{-1+\lambda} e^{\alpha Q(s,a)} \xrightarrow{\text{قرار دادن در شرط}} e^{-1+\lambda} \sum_a e^{\alpha Q(s,a)} = 1$$

$$\Rightarrow e^{1-\lambda} = \frac{1}{\sum_a e^{\alpha Q(s,a)}} \quad \rightsquigarrow p(a|s) = \frac{e^{\alpha Q(s,a)}}{\sum_{a'} e^{\alpha Q(s,a')}}$$

در نتیجه مسأله کلی به شکل زیر تبدیل می‌شود:

$$\min_Q \alpha E_{s\sim D}\Big[\sum_a p(a|s)\big(Q(s,a) - Q(s,a) + \log \sum_a e^{Q(s,a)}\big) - E_{a\sim\hat{\pi}_B}[Q(s,a)]\Big]$$
$$+ \frac{1}{2} E_{s,a,s'}\big[(Q - \hat{B}^{\pi_k}\hat{Q}^k)^2\big]$$

$$\rightsquigarrow \min_Q \alpha E_{s\sim D}\Big[\log \sum_a e^{Q(s,a)} - E_{a\sim\hat{\pi}_B}[Q(s,a)]\Big] + \frac{1}{2} E_{s,a,s'}\big[(Q - \hat{B}^{\pi_k}\hat{Q}^k)^2\big]$$

پس به کمک این روش می‌توانیم مسأله داخلی را حل کنیم و فرم کلی را ساده کنیم. همچنین این کار مانع از تخمین زیاد مقدارها روی Q های بزرگ می‌شد، آن را کاهش می‌دهد و اکسپلوریشن به مسأله ما اضافه می‌شود.

conservative polic

ج) مسأله به شکل زیر تبدیل می‌شود:

$$\pi^* = \arg\max_{\pi} E_{s\sim D, a\sim\pi(a|s)}[Q(s,a)] + R(\pi)$$

$$\min_{Q} \alpha E_{s\sim D, a\sim\pi^*(a|s)}[Q(s,a)] + \frac{1}{2} E_{s,a,s'\sim D}[(Q - \hat{B}^{\pi_k}\hat{Q}^k)^2] + R(\pi^*)$$

برای هر s مسأله را جداجدا حل می‌کنیم:

$$\min_{Q} \alpha E_{a\sim\pi^*(a|s)}[Q(s,a)] + \frac{1}{2}\sum_a \hat{\pi}_\beta(a|s)(Q(s,a) - \hat{B}^{\pi_k}\hat{Q}^k)^2$$

مشتق مسأله را برابر 0 قرار می‌دهیم و داریم:

$$\alpha\pi^*(a|s) + \hat{\pi}(a|s)(\hat{Q}^{k+1}(s,a) - \hat{B}^{\pi_k}\hat{Q}^k(s,a)) = 0$$

$$\rightsquigarrow \hat{Q}^{k+1}(s,a) = \hat{B}^{\pi_k}\hat{Q}^k(s,a) - \alpha\frac{\pi^*(a|s)}{\hat{\pi}_\beta(a|s)}$$

د) با توجه به نقطه ثابت مدل و کران به دست آمده داریم (برای fixed point):

$$\hat{Q}^\pi(s,a) = \hat{B}^\pi\hat{Q}^\pi(s,a) - \alpha\frac{\pi^*(a|s)}{\hat{\pi}_\beta(a|s)} \xrightarrow{-C_\delta(s,a) - \hat{B}^\pi\hat{Q}(s,a) - \hat{B}^\pi C_\delta(s,a)}$$

$$\hat{Q}^\pi(s,a) \le \hat{B}^\pi\hat{Q}(s,a) + C_\delta(s,a) - \alpha\frac{\pi^*(a|s)}{\hat{\pi}_\beta(a|s)} = R_{(s,a)} + \gamma P^\pi\hat{Q}_{(s,a)} + C_\delta(s,a) - \alpha\frac{\pi^*(a|s)}{\hat{\pi}_\beta(a|s)}$$

$$\rightarrow (I - \gamma P^\pi)\hat{Q}^\pi(s,a) \le R(s,a) + C_\delta(s,a) - \alpha\frac{\pi^*(a|s)}{\hat{\pi}_\beta(a|s)}$$

$$\rightarrow \hat{Q}^\pi(s,a) \le \underbrace{(I - \gamma P^\pi)^{-1}R(s,a)}_{Q^\pi(s,a)} + \frac{1}{\delta}(I - \gamma P^\pi)^{-1}C_\delta(s,a) - (I - \gamma P^\pi)^{-1}\alpha\frac{\pi^*(a|s)}{\hat{\pi}_\beta(a|s)}$$

$$\rightarrow \hat{Q}^\pi(s,a) \le Q^\pi(s,a) + (I - \gamma P^\pi)^{-1}C_\delta(s,a) - \alpha(I - \gamma P^\pi)^{-1}\frac{\pi^*(a|s)}{\hat{\pi}_\beta(a|s)}$$

Subject: Date: / /

2) الف) BC سعی می کند پالیسی ها را به صورت مستقیم از اکشن های متخصص یاد بگیرد که این باعث می شود که توانایی generalization آن کاهش یابد و بسیار به کیفیت داده ها حساس شود. IRL سعی می کند ابتدا ریوارد فانکشن را به دست آورد که این سبب می شود تا قدرت تعمیم ما بالاتر رود و نسبت به نویزها مقاوم تر شود. این باعث می شود تا روش های IRL انعطاف پذیرتر و سازگارتر شوند و انگیزه های پشت رفتارهای expert را درک کند نه اینکه صرفا حرکات را تقلید کند. همچنین IRL می تواند پالیسی های sub optimal را بهتر کند.

ب) در این مسائل مانند ترم آنتروپی اضافه می کنیم و آن را ماکسیمایز می کنیم تا ابهام را رفع کنیم. این باعث می شود تا از بین تمام پالیسی های ممکن که می تواند رفتار expert را توجیه کند آن را برداریم که بیشترین آنتروپی و به طبع بیشترین رندومنس را دارد. در این حالت مطمئن می شویم که پالیسی ما تا حد امکان رندوم است و همچنان مطابق رفتار expert است که این از overfit شدن جلوگیری می کند. همچنین exploration را تقویت می کند و باعث مقاوم شدن آن نسبت به variation های مختلف دیتا می شود. سبب افزایش generalization

$$\max_{\theta} \sum_{\tau\sim D} \log p(\tau|\theta) - \lambda \| E_{\pi_\theta}[\phi(s,a)] - E_{\pi_E}[\phi(s,a)] \| \qquad p(\tau|\pi) = \frac{1}{Z(\theta)} \exp\left(\sum_{t=0}^{T} R(s_t,a_t)\right)$$

ج) $J(\pi, f) = \lambda H(\pi) - E_{\pi}[\log(D(s,a))]$ $\qquad J(\pi_E, f) = E_{\pi_E}[\log(1-D(s,a))]$

D تلاش می کند تا فاصله بین پالیسی expert و پالیسی در حال یادگیری را افزایش دهد. در حقیقت تلاش می کند [illegible] مدلی را خروجی دهد که در آن پالیسی در حال یادگیری در بدترین حالت خود نسبت به پالیسی expert باشد.

s.a.m

در این حالت با بیشینه کردن نسبت به π داریم پالیسی مان را در بدترین حالت ممکن بهبود می بخشیم که این همان نقش adverserial آن است. حال با بیشینه کردن $H(\pi)$ یا همان ماکسیمایز کردن $H(\pi)$ در حقیقت داریم تضمین می کنیم که پالیسی تا جای ممکن رندوم است و شکل نقش ب را کامل می کنیم.

د) روش های مبتنی بر ویژگی مانند maxEnt IRL با مشکلاتی مانند نیاز به مهندسی ویژگی گسترده و محدودیت در انتخاب ویژگی ها مواجه اند مشکلات scalability و generalization نسبت به داده های دیده نشده را دارند. GAIL با استفاده از شبکه های عصبی عمیق و یادگیری ویژگی ها از داده های خام state-action این مشکلات را برطرف می کند و نیاز به داده های دست ساز اضافه را کاهش می دهد. این باعث افزایش مقیاس پذیری و تعمیم می شود زیرا چارچوب adverserial روش GAIL تقلید جامع تری از expert را فراهم می کند و محدودیت های روش های قبلی را برطرف می کند. یادگیری adverserial همزمان هم سیاست و هم discriminator را بهبود می بخشد.

Subject: Date: / /

3) الف) روش هایی که تنها از سمپل های model-based استفاده می کنند، اغلب از عدم دقت مدل رنج می برند که می تواند به سیاست های sub optimal منجر شود. از سوی دیگر روش هایی که فاصله تعاملات دنیای واقعی

مثلن اند می تواند پرهزینه، زمانبر و خطرناک باشد به خصوص در سمپل هایی که پیچیده یا ناامن هستند. روش های ترکیبی

مانند MBPO نقاط قوت هر دو رویکرد را با هم ترکیب می کند. آنها از شبیه سازی های بیشتر مدل برای ترکیب داده های

آموزش اضافی استفاده می کنند که sample efficiency را افزایش می دهد و همچنین تعاملات دنیای واقعی را برای

تصحیح بایاس و نواقص مدل به کار می برند که منجر به سیاست های مقاوم تر و قوی تر می شود. این رویکرد ترکیبی

تعادل ایجاد می کند که از طراحی مدل ها استفاده می کند در حالی که با بازخورد واقعی، قابلیت اطمینان را تضمین می کند.

ب) MBPO شبیه سازی بیشتر مدل را با استفاده از مدل آموزش دیده برای ترکیب rollout ها ترکیب می کند

که سپس در مدل شبیه سازی با داده های دنیای واقعی ترکیب می شوند. trade off مدل های مختلف برای rollout

ها همان تعادل بین بایاس و واریانس است. مدل کوتاه سبب خطاهای تعمیم می شود اما علن است

نتواند به اندازه کافی explore کند که سبب افزایش بایاس و کاهش واریانس می شود. مدل بیشتر سبب می شود

تا داده های بیشتری را مشاهده کنیم و این خطاهای تعمیم مدل را افزایش می دهد و سبب کاهش دقت مدل می شود.

MBPO به طور بهینه خود را تطبیق می دهد تا از مزایای هر دو به صورت همزمان استفاده کند. سپس معمولا 3 تا مرحله

s.a.m

ج) MBPO نمی‌تواند در محیط‌های آفلاین استفاده شود زیرا نیاز به تعامل با محیط واقعی دارد و در یادگیری آفلاین ما این دسترسی را نداریم. این باعث می‌شود خطاهای مدل ما به صورت فزاینده ای افزایش یابد و منجر به عملکرد ضعیف سیاست شود و نتوانیم مشکلات OOD را حل کنیم.

د) یکی از روش‌های عملی برای اندازه‌گیری عدم قطعیت به این صورت است که از مدل‌های انسامبل استفاده کنیم. در این حالت چند مدل روی یک دیتا ترین می‌شود و نتایج آنها با هم ترکیب و مقایسه می‌شود. واریانس بین این مدل‌ها به ما میزان عدم قطعیت را می‌دهد. پس سیاست‌ها از ورود به مناطق با عدم قطعیت بالا منصرف می‌شوند و عملکرد ایمن‌تر و قابل اعتمادتری تضمین می‌شود.

ه) COMBO استراتژی‌های محافظه‌کارانه را با ایده‌های Dyna و CQL ترکیب می‌کند تا چالش‌های یادگیری آفلاین را حل کند. روش Dyna از مدل برای شبیه‌سازی تجربه‌ها برای آپدیت کردن پالیسی استفاده می‌کند و با روش CQL که به صورت conservative پالیسی اپتیمایزیشن انجام می‌دهد ادغام می‌شود. (اساساً از نقاط پالیسی دور شده و دیتاست بهینه‌تر را جریمه می‌کند). این ترکیب اطمینان می‌دهد که مدل مناطق باقی می‌باشد دارای over estimate ریسک‌ها در مناطق با داده ناکافی جلوگیری می‌کند و تجربیات شبیه‌سازی شده را برای بهبود استحکام و عملکرد سیاست به کار می‌گیرد. این روش ترکیبی به طور مؤثر تعادل بین اکتشاف و محافظه‌کاری را برای بهبود نتایج یادگیری فراهم می‌کند.

$$Q_{new}(s,a) = \min\{Q_{old}(s,a),\ r + \gamma E_{a'\sim\pi}[Q_{target}(s',a')]\}$$

د) مدل های محافظه کارانه مانند COMBO و MOREL، ایمنی و robustness را با جلوگیری از اکشن های ریسکی و جریمه کردن استیت های غیر مطمئن اولویت بندی می کنند و برای این کار از روش های متفاوتی استفاده می کنند.

COMBO از آپدیت پالیسی محافظه کارانه استفاده می کند تا از اکشن های safe مطمئن شود در حالی که MoRel به صورت خاص ناحیه ها با عدم قطعیت بالا را جریمه می کند

COMBO ـــ نقاط قوت: ۱) الگوریتم متعادل (محافظه کارانه و فعال) 2) sample efficient 3) جلوگیری از over estimation ارزش اکشن ها و در نتیجه جلوگیری از اکشن های ریسکی که سبب پرفورمنس ضعیف می شوند

نقاط ضعف: ۱) امکان محافظه کارانه بودن بیش از حد که سبب محدود کردن پالیسی می شود 2) پیچیدگی پیاده سازی

MoRel ـــ نقاط قوت: ۱) مدل کردن عدم قطعیت به صورت مستقیم و صریح 2) تضمین ایمنی بالا

نقاط ضعف: ۱) جستجوی محافظه کارانه که ممکن است به جواب sub optimal در مناطق کمتر ریسکی ختم شود

2) زمانبر بودن به خاطر استفاده از انسامبل 3) ممکن است فرصت پیدا کردن مناطق با ریوارد بالا را از دست بدهد

اثبات‌ها با کمک کتاب الگوریتم‌های بندیت Csaba Szepesvári, Tor Lattimore

4) الف) نابرابری Hoeffding را می‌نویسیم،

$$S_n = \sum_{i=1}^{n} X_i \qquad a_i \le X_i \le b_i \qquad \bar{X}_n = \frac{S_n}{n}$$

$$P(E[S_n] - S_n \ge t) \le \exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right) \qquad X_i \to i.i.d$$

در صورتی که $b_i = 1$ و $a_i = 0$ باشد داریم:

$$P(E[X] - \bar{X}_n \ge \frac{t}{n}) \le \exp\left(\frac{-2t^2}{4n}\right)$$

$$\Rightarrow P(E[X] - \bar{X}_n \ge \varepsilon) \le \exp\left(\frac{-n\varepsilon^2}{2}\right) \Rightarrow P(\mu \ge \hat{\mu} + \varepsilon) \le \exp\left(\frac{-n\varepsilon^2}{2}\right)$$

در جای ε مقدار دهیم $\sqrt{\frac{2\log\frac{1}{\delta}}{n}}$ داریم،

$$P\left(\mu \ge \hat{\mu} + \sqrt{\frac{2\log\frac{1}{\delta}}{n}}\right) \le \exp\left(\frac{-n\frac{2\log\frac{1}{\delta}}{n}}{2}\right) = \delta$$

$$\Rightarrow P\left(\mu \ge \hat{\mu} + \sqrt{\frac{2\log\frac{1}{\delta}}{n}}\right) \le \delta$$

در نتیجه با احتمال حداقل $1-\delta$ داریم که:

$$\mu \le \hat{\mu} + \sqrt{\frac{2\log\frac{1}{\delta}}{n}}$$

و این بالا برای میانگین پاداش هر عمل است.

ب) فرض می‌کنیم K بازو و n گام زمانی داریم و بدون کم شدن از کلیت مسأله در نظر می‌گیریم که بازوی اول بهینه است.

یعنی $\mu_1 = \mu^*$. فرض می‌کنیم پاداش مشاهده در t امین اجرای بازوی i را با X_{t_i} نشان می‌دهیم و همچنین میانگین تجربی پاداش حاصل از انجام آن n بار اجرای آن را با $\hat{\mu}_{im}$ نشان می‌دهیم که برابر است با $\frac{\sum_{t=1}^{m} X_{t_i}}{m}$

برای اینکه بتوانیم برای $E[T_i(n)]$ به دست آوریم ابتدا باید رویداد G_i را به شکل زیر تعریف کنیم،

$$G_i = \left\{\mu_1 < \min_{t\in[n]} UCB_1(t,\delta)\right\} \cap \left\{\hat{\mu}_{iu_i} < \sqrt{\frac{2\log\frac{1}{\delta}}{u_i}} + \mu_1\right\}$$

s.a.m

دو مورد را اثبات می‌کنیم: ① در صورتی که G_i رخ دهد، بازوی i حداقل u_i بار انتخاب شده است.

* اثبات: از برهان خلف استفاده می‌کنیم. فرض می‌کنیم $T_i(n) < u_i$ باشد. در این صورت به دلیل اینکه بازوی i بیشتر از u_i بار انتخاب شده است، باید یک $t \in [n]$ وجود داشته باشد که $A_t = i$ و $T_i(t-1) = u_i$. در نتیجه داریم:

$$UCB_i(t-1,\delta) = \hat{\mu}_i(t-1) + \sqrt{\frac{2\log\frac{1}{\delta}}{T_i(t-1)}} = \hat{\mu}_{i,u_i} + \sqrt{\frac{2\log\frac{1}{\delta}}{u_i}} \overset{?}{<} \mu_i < UCB_i(t-1,\delta)$$

$$\Rightarrow A_t = \operatorname{argmax}_j UCB_j(t-1,\delta) \neq i \quad \text{✗} \qquad \checkmark$$

② G_i با احتمال زیاد رخ می‌دهد.

* اثبات: این به این معنیست که G_i^c به احتمال کمی رخ می‌دهد. ضمن تعریف داریم:

$$G_i^c = \{\mu_i \geq \min_{t\in[n]} UCB_i(t,\delta)\} \cup \{\hat{\mu}_{i,u_i} + \sqrt{\frac{2\log\frac{1}{\delta}}{u_i}} \geq \mu_i\}$$

ضمن union bound:

$$P(\mu_i \geq \min_{t\in[n]} UCB_i(t,\delta)) = P(\bigcup_{t\in[n]} \mu_i \geq \hat{\mu}_{i,T_i(t)} + \sqrt{\frac{2\log\frac{1}{\delta}}{T_i(t)}}) \overset{T_i(t)\leq t}{\leq} P(\bigcup_{t\in[n]} \mu_i \geq \hat{\mu}_{it} + \sqrt{\frac{2\log(\frac{1}{\delta})}{t}})$$

$$\leq \sum_{t=1}^{n} P(\mu_i \geq \hat{\mu}_{it} + \sqrt{\frac{2\log(\frac{1}{\delta})}{t}}) \overset{\text{هافدینگ}}{\leq} n\delta$$

حال فرض می‌کنیم بازوی i را به صورتی انتخاب می‌کنیم که (ضمن صورت سوال $\mu_1 > \mu_i$، $\Delta_i = \mu_1 - \mu_i$):

$$\frac{1}{2}\Delta_i \geq \sqrt{\frac{2\log\frac{1}{\delta}}{u_i}} \quad *$$

$$\Rightarrow P(\hat{\mu}_{i,u_i} + \sqrt{\frac{2\log\frac{1}{\delta}}{u_i}} \geq \mu_i) = P(\hat{\mu}_{i,u_i} - \mu_i \geq \Delta_i - \sqrt{\frac{2\log\frac{1}{\delta}}{u_i}}) \overset{*}{\leq} P(\hat{\mu}_{i,u_i} - \mu_i \geq \frac{1}{2}\Delta_i) \overset{\text{هافدینگ}}{\leq} \exp(\frac{-u_i\Delta_i^2}{8})$$

در نتیجه:

$$P(G_i^c) \leq n\delta + \exp(\frac{-u_i\Delta_i^2}{8}) \qquad \checkmark$$

s.a.m

Subject: Date: / /

حال صورت سوال را ثابت می‌کنیم. برای $E[T_i(n)]$ داریم:

$$E[T_i(n)] \le \underbrace{E[I(G_i)T_i(n)]}_{\le u_i} + \underbrace{E[I(G_i^c)T_i(n)]}_{\le np(G_i^c)} \le u_i + np(G_i^c) \le u_i + n\left(n\delta + \exp\left(-\frac{u_i\Delta_i^2}{8}\right)\right)$$

اگر u_i را برابر $\left\lceil \frac{8\log\frac{1}{\delta}}{\Delta_i^2} \right\rceil$ قرار دهیم (دقت کنید که در * صدق می‌کند) داریم:

$$E[T_i(n)] \le \left\lceil \frac{16\log n}{\Delta_i^2} \right\rceil + n(n\delta+\delta) \underbrace{\delta = \frac{1}{n^2}}_{1+\frac{1}{n} \le 2} \le \frac{16\log n}{\Delta_i^2} + 1 + 2 = \frac{16\log n}{\Delta_i^2} + 3 \quad \checkmark$$

ج) ابتدا اثبات می‌کنیم که R_n از رابطه روبرو به دست می‌آید:

$$R_n = \sum_{i=1}^{K} \Delta_i E[T_i(n)]$$

* ابتدا طبق تعریف، regret داریم:

$$R_n = n\mu^* - E\left[\sum_{t=1}^{n} X_t\right]$$

و در هر اندیس داریم $\sum_{i=1}^{K} I\{A_t = i\} = 1$. در نتیجه:

$$R_n = n\mu^* - \sum_{t=1}^{n}\sum_{i=1}^{K} E[I\{A_t=i\}X_t] = \sum_{i=1}^{K}\sum_{t=1}^{n} E[(\mu^* - X_t)I\{A_t=i\}]$$

حال طبق قانون امید ریاضی شرطی داریم:

$$= \sum_{i=1}^{K}\sum_{t=1}^{n} E[E[(\mu^* - X_t)I\{A_t=i\}|A_t]]$$

$$\Rightarrow = \sum_{i=1}^{K}\sum_{t=1}^{n} E[I\{A_t=i\}\underbrace{E[(\mu^* - X_t)|A_t]}_{\mu^* - \mu_{A_t} = \Delta_i}] = \sum_{i=1}^{K}\sum_{t=1}^{n} \Delta_i E[I\{A_t=i\}] = \sum_{i=1}^{K} \Delta_i \sum_{t=1}^{n} E[I\{A_t=i\}]$$

$$\Rightarrow = \sum_{i=1}^{K} \Delta_i E[T_i(n)] \quad \checkmark$$

از طرفی حق نیز داریم:

$$E[T_i(n)] \le \frac{16\log n}{\Delta_i^2} + 3$$

در نتیجه:

$$R_n = \sum_{i=1}^{K} \Delta_i E[T_i(n)] \le \sum_{i=1}^{K} 3\Delta_i + \sum_{i:\Delta_i>0} \frac{16\log n}{\Delta_i} \quad \checkmark$$

s.a.m

Subject: Date: / /

$$R_n = \sum_{i=1}^{K} \Delta_i E[T_i(n)]$$

د) اثبات دقت مثل:

این حد را به دو بخش برای دلتاهای کوچکتر و بزرگتر مساوی $\Delta = \sqrt{\frac{16K\log n}{n}}$ تقسیم می‌کنیم:

$$R_n = \sum_{i:\Delta_i<\Delta} \Delta_i \overbrace{E[T_i(n)]}^{<n} + \sum_{i:\Delta_i\geq\Delta} \Delta_i E[T_i(n)] \leq n\Delta + \sum_{i:\Delta_i\geq\Delta} \left(3\Delta_i + \frac{16\log n}{\Delta_i}\right)$$

(ضرب)

$$\leq n\Delta + \sum_{i=1}^{K} 3\Delta_i + \frac{16K\log n}{\Delta} = \sqrt{16Kn\log n} + \sqrt{16Kn\log n} + \sum_{i=1}^{K} 3\Delta_i = 8\sqrt{Kn\log n} + 3\sum_{i=1}^{K} 3\Delta_i$$

s.a.m