شکل‌دهی موثر عقیده: رویکرد یادگیری تقویتی عمیق در تعاملات ربات و کاربر

دانشجو: **صبا صمدی**

استاد راهنما: **دکتر حامد کبریایی**

دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران

مقدمه

با گسترش پلتفرم‌های رسانه‌های اجتماعی، نظرات افراد در رابطه با یک موضوع خاص، در معرض نظرات افراد دیگر قرار می‌گیرد که دیدگاه‌ها و تصمیمات آنها را تغییر می‌دهد. این امر که به پویایی عقیده معروف است در پیش‌بینی و شکل‌دهی نظرات نقش حائز اهمیتی دارد. در این پروژه دو سناریوی شکل‌دهی عقیده با استفاده از یادگیری تقویتی عمیق در محیط مدل اطمینان محدود تصادفی (SBCM) ارائه و بررسی شده است:

- الگوریتم SBCM در حضور ربات با نظرات کنترل شده
- الگوریم SBCM در حضور تبلیغات و تحت بودجه مشخص

روش پیشنهادی

- **الگوریتم Stochastic Bounded Confidence Model**

الگوریتم‌های اطمینان محدود شده، همواره نظر افراد را با توجه به نظرات افراد حول آن فرد به‌روزرسانی می‌کند.

$$x_u(t+1) = x_u(t) + \frac{1}{|\mathcal{N}_u(t)|} \sum_{v \in \mathcal{N}_u(t)} (x_v(t) - x_u(t))$$

در رویکرد اطمینان محدود شده تصادفی، یک توزیع احتمالاتی برای هر فرد ایجاد می‌گردد که به میزان اختلاف نظرات فرد با بقیه افراد وابسته است. سپس افراد به جهت تعامل، بر اساس این توزیع احتمال نمونه‌برداری می‌شوند.

$$p(z_{uv}^t = 1) = \frac{|x_u(t) - x_v(t)|^{-\epsilon}}{\sum_{v'} |x_u(t) - x_{v'}(t)|^{-\epsilon}}$$

- **سناریوی تعامل بات-کاربر در محیط SBCM**

در هر گام زمانی، عملگر نظرات ربات‌های تحت کنترل را بسته به نظرات کاربران تنظیم می‌کند. سپس نظرات کاربران در محیط SBCM به روز رسانی می‌شود که این امر تا پایان اپیزود تکرار می‌شود. در این شبیه‌سازی، r_t (پاداش در زمان t) اختلاف میانگین نظرات کاربران، تعریف گردیده است.

$$r_t = (\frac{t}{20}) \times [\mu(u_t) - \mu(u_{t-1})]$$

Algorithm 1 User-Bot Interaction in SBCM

procedure DDPG_TRAIN(N, T)
  Initialize Actor: $\pi(s;\theta^\pi)$ and $\pi'(s;\theta^{\pi'})$
  Initialize Critic: $Q(s,a;\theta^Q)$ and $Q'(s,a;\theta^{Q'})$
  Initialize Replay Buffer $\mathcal{R}$
  Initialize noise process $\mathcal{N}$
  for episode = 1 to M **do**
    Reset noise process: $\mathcal{N}$.reset()
    Sample initial user opinions: $u_0 \sim U(-1,1)$
    for t = 1 to T **do**
      Augment state with time: $s_{t-1} = [u_{t-1}, t]$
      Select action: $a_{t-1} = \pi(s_{t-1}) + \mathcal{N}$.sample()
      $u_t, r_t \leftarrow \text{SBCM}(N, u_{t-1}, a_{t-1})$
      Store $(s_{t-1}, a_{t-1}, r_t, s_t)$ in $\mathcal{R}$
      if size of $\mathcal{R} \geq$ batch_size **then**
        Update Q and π using mini-batch from $\mathcal{R}$
        Softly update π' and Q' with τ

شبه کد ۱: شبه کد الگوریتم مورد استفاده

- **سناریوی تبلیغ تحت بودجه مشخص در محیط SBCM**

در هر گام زمانی، عملگر نظرات کاربران، گام زمان کدگذاری‌شده و بودجه باقی‌مانده را دریافت میکند و دو مقدار مکان و محدوده تبلیغات را به عنوان خروجی میدهد.

$$r_t = \begin{cases} (\frac{t}{20}) \times [\mu(u_t) - \mu(u_{t-1})], & \text{if } B_t \geq 0 \\ (\frac{t}{20}) \times [\mu(u_t) - \mu(u_{t-1})] - C_t, & \text{if } B_t < 0 \end{cases}$$

$$Cost_{\text{range}} = C_r \times \left(e^{(\text{spread} \times p)} - 1\right)$$

$$Cost_{\text{opinion}} = C_o \times \left(e^{(|\text{location}| \times q)} - 1\right)$$

$$C_t = Cost_{\text{range}} + Cost_{\text{opinion}}$$

Algorithm 2 Advertising in SBCM

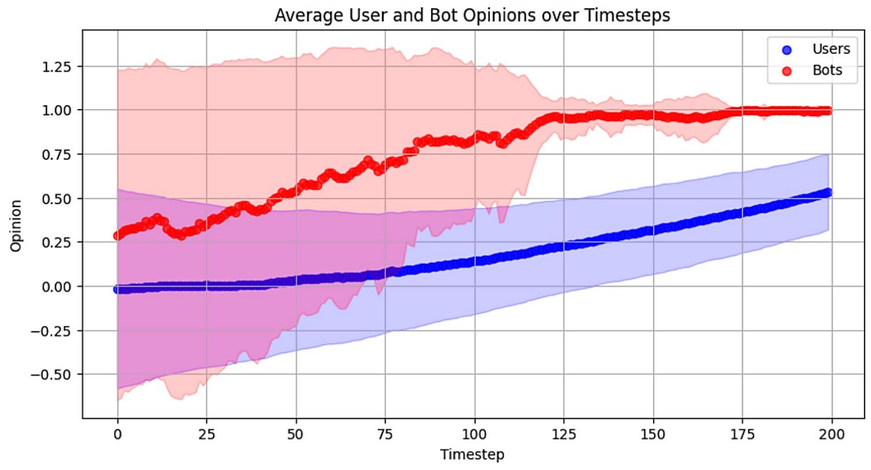
procedure DDPG_TRAIN(N, T, B_0)
  Initialize Actor: $\pi(s;\theta^\pi)$ and $\pi'(s;\theta^{\pi'})$
  Initialize Critic: $Q(s,a;\theta^Q)$ and $Q'(s,a;\theta^{Q'})$
  Initialize Replay Buffer $\mathcal{R}$
  Initialize noise process $\mathcal{N}$
  for episode = 1 to M **do**
    Reset noise process: $\mathcal{N}$.reset()
    Sample initial user opinions: $u_0 \sim U(-1,1)$
    Initialize remaining budget: $B = B_0$
    for t = 1 to T **do**
      State Representation: $s_{t-1} = [u_{t-1}, t, B]$
      Select action: $a_{t-1} = \pi(s_{t-1}) + \mathcal{N}$.sample()
      $(u'_{t-1}, C_t) \leftarrow \text{Advertisement}(u_{t-1}, a_{t-1})$
      $B \leftarrow B - C_t$
      $u_t, r_t \leftarrow \text{SBCM}(N, u'_{t-1}, a_{t-1})$
      if $B < 0$ **then**
        $r_t \leftarrow r_t - C_t$
      Store $(s_{t-1}, a_{t-1}, r_t, u_t, B)$ in $\mathcal{R}$
      if size of $\mathcal{R} \geq$ batch_size **then**
        Update Q and π using mini-batch from $\mathcal{R}$
        Softly update π' and Q' with τ

شبه کد ۲: شبه کد الگوریتم مورد استفاده

نتایج

برای تجزیه و تحلیل، فرض شده است که تعداد کاربران و واحد زمانی برابر ۲۰۰ میباشد.

- **ارزیابی سناریوی تعامل ربات-کاربر**

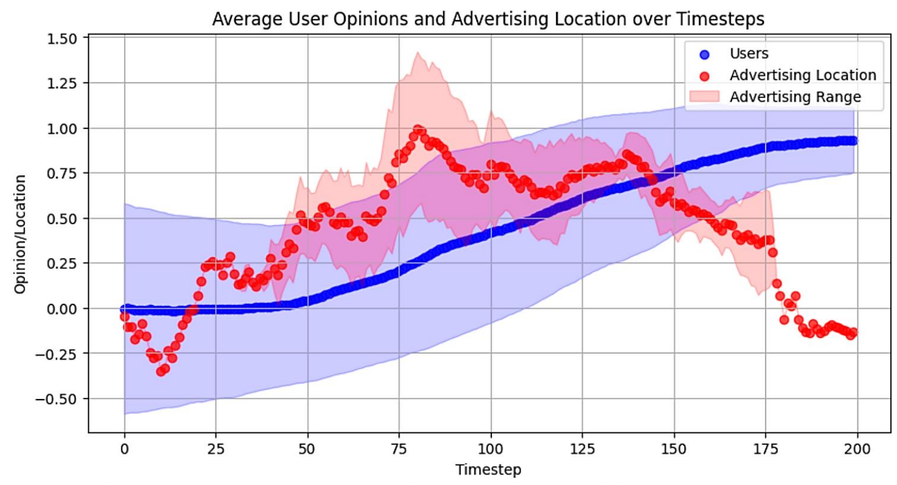


شکل ۱: نحوه تکامل نظرات کاربران و ربات‌ها در یک اپیزود

Number of Bots	$\mu = 0.1, \epsilon = -2.0$		$\mu = 0.05, \epsilon = -1.0$	
	Mean	STD	Mean	STD
20	0.556	0.071	0.573	0.074
15	0.481	0.075	0.117	0.074
10	0.389	0.083	0.074	0.072

جدول ۱: ارزیابی میانگین و انحراف معیار نظرات در سناریوی کاربر-ربات برای ۱۰۰۰ قسمت

- **ارزیابی سناریوی تبلیغات تحت بودجه مشخص**



شکل ۲: نحوه تکامل نظرات کاربران و نحوه انجام تبلیغات هدفمند در یک اپیزود

Budget	$\mu = 0.1, \epsilon = -2.0$		$\mu = 0.05, \epsilon = -1.0$	
	Mean	STD	Mean	STD
20	0.851	0.190	0.321	0.433
10	0.735	0.214	0.267	0.423
5	0.617	0.224	0.226	0.414

جدول ۲ - ارزیابی میانگین و انحراف معیار نظرات در سناریوی تبلیغات

جمع بندی

در این مقاله، ارائه شده است که چگونه در SBCM با استفاده از الگوریتم DDPG شکل‌گیری نظرات انجام شود. ابتکار مهم این بود که عوامل هوش مصنوعی که یک گروه از ربات‌ها را کنترل می‌کنند، به شبکه اضافه شوند. ربات‌ها که می‌توانند به دلخواه نظرات خود را تنظیم کنند و در برابر نظرات کاربران مصون هستند، برای هدایت نظر میانگین کاربران به سمت یک جهت مورد نظر استفاده شدند. یک استراتژی جایگزین شامل استخدام یک عامل تبلیغاتی است که با در نظر گرفتن یک بودجه تبلیغاتی مشخص، استراتژی‌های بهینه برای مدیریت نظرات براساس مشاهدات خود یاد می‌گیرد. نتایج آزمایشی اثربخشی و قابلیت مقیاس پذیری الگوریتم‌های پیشنهادی را تایید می‌کنند. این مطالعه مسائل مهمی در خصوص استفاده اخلاقی از هوش مصنوعی در شبکه‌های اجتماعی و پلتفرم‌های انتشار اطلاعات را مطرح می‌کند.

مراجع اصلی

[1] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. part ii," *Annual Reviews in Control*, vol. 45, pp. 166–190, 2018.

[2] M. Okawa and T. Iwata, "Predicting opinion dynamics via sociologically-informed neural networks," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2022.

[3] F. Baccelli, A. Chatterjee, and S. Vishwanath, "Stochastic bounded confidence opinion dynamics," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 3408–3413.

[4] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2019.