用于大语言模型时间情感模式识别的连续时间建模

Rezky M. Kam计算机科学学院 *Bina Nusantara*大学
东雅加达，印度尼西亚
rezky.kam@binus.ac.id

Coddy N. Siswanto计算机科学学院 *Bina Nusantara*大学
东雅加达，印度尼西亚
coddy.siswanto@binus.ac.id

arXiv:2601.12341v1 [cs.LG] 18 Jan 2026

摘要— **文本模态解码器模型依赖离散令牌生成，前向注意力权重会影响单个令牌表示，以根据反向传播过程中学习到的模式预测最可能的输出。然而，仅依靠这种方法往往缺乏真正的理解，并且会将对话中的情感动态视为黑箱；真正的情感理解涉及基于轮次的对话中不断演变的情感轨迹的同步动态。此外，由于训练数据存在差异，包括混合专家解码器中包含大量专家，因此无法通过启发式度量来量化判断其属于哪种情感人格。在资源受限的环境中，在线强化学习的主流应用也面临挑战。此外，我们提出一种混合编码器—解码器架构，利用解码器渐进式引导情感轨迹。具体而言，该架构具有时间感知能力，基于连续时间模式及其与聚焦群体领域情感状态变化性的相关性，并借助物理信息神经网络实现时间适应与纵向适应。我们的方法使机器能够在交互过程中模仿用户的心理学合理性，同时保留独立的表达能力、计算效率和可解释性。**

I. 引言

解码器文本生成模型可以通过选择概率最高的令牌，识别单次对话交流中的复杂语义含义和情绪语境 [1]，[2]。尽管从更广泛的技术视角来看，由于内部机制不透明，尤其是这些模型如何处理在大量且多样化语料库中训练得到的情感状态变化性，它们通常被视为黑箱。解码器微调的一种直接替代方案是基于提示的学习，例如使用微调编码器进行零样本或少样本提示 [3], [4]。在这些方法中，在多标签情感提取任务中，编码器的输出通常通过原始逻辑值或温度缩放逻辑值进行特征工程处理：

$$p = \{e_i : \tilde{z} | i \in [0, n-1]\}$$

$$p(e_i) = \frac{\exp(\tilde{z}_i)}{\sum_{j=0}^{n-1} \exp(\tilde{z}_j)} = \frac{\exp(z_i/\delta)}{\sum_{j=0}^{n-1} \exp(z_j/\delta)}$$

此处，缩放因子 δ 在应用 Softmax 后调节最终的 logit 分布p。当 $\delta \to 0$时，该分布趋近于独热编码，通常用于提取置信度最高的预测：

$$\text{Prompt}(\max(p(e_i))) \to \text{Predict}(\text{tokens}(p(e_i)))$$

相反，当 $\delta \to \infty$ 时，分布会变得均匀，而置信度较低的输出则可用于引入多样性或细腻的理解：

$$\text{Prompt}(p(e_{i \neq i^*})) \to \text{Predict}(\text{tokens}(p(e_{i \neq i^*})))$$

这些策略常用于领域特定聊天机器人，尤其是在陪伴、心理健康支持和情感计算等应用中。然而，基于令牌的生成具有离散且静态的特性，这构成了一项根本性局限。相比之下，人类情绪本质上是连续的，并会随着时间推移，通过动态的情感轨迹不断演变 [5]。因此，当前的语言模型难以捕捉情感过程在时间维度和心理层面的真实性，导致其适应性和可解释性有限。

为解决这一局限，我们提出了一种编码器—解码器框架，将连续时间建模的时间敏感性与大型语言模型的表征能力相结合；此外，CEmoFlow数据集包含每条话语的情绪软标签、时间戳和延迟。我们不再仅通过提示工程或数据驱动的微调来增强情感理解，而是利用由时间感知微分系统调制的上下文向量（ICVs） [6]，，直接拦截文本模态解码器的前向传播。具体而言，我们利用在连续情感数据集 [9],上训练的连续时间循环神经网络（CTRNNs） [7] 和神经常微分方程（Neural ODEs） [8]，，引导令牌生成沿着心理学上合理的情感轨迹进行。

这一架构使解码器能够在多轮交互过程中捕捉情感状态细粒度的纵向变化。此外，通过将物理信息神经网络融入情感建模流程，我们引入领域知识来规范时间动态，从而提升泛化能力和可解释性。重要的是，我们的方法在增强生成的对话的情感保真度的同时，保留了自回归生成的计算效率。

通过这种方式，我们弥合了符号离散建模与亚符号连续情感表征之间的差距，使对话代理能够更具表现力、情感上更加连贯，并更贴合用户需求。这项贡献不仅对情感人工智能和人机交互具有意义，也有助于推动更广泛的可解释人工智能和以心理学为基础的人工智能研究。

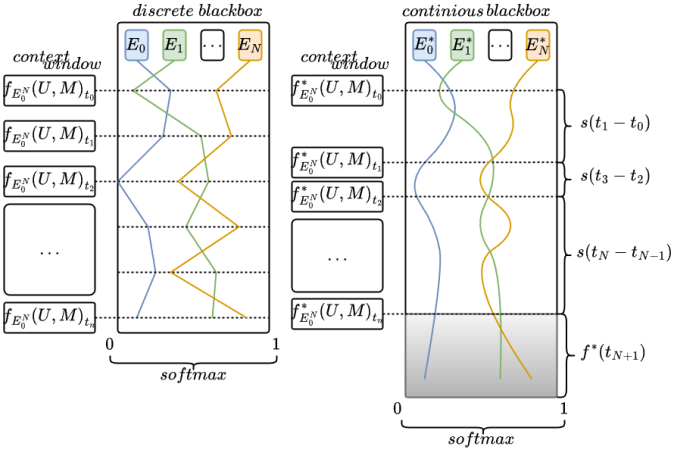

图1。令 E 两个函数的 $f = encoder_{out(E_0^N)}$ 以离散形式呈现，而 $f^* = encoder \rightarrow LatentODE_{out(E_0^N)}$ 以 连续 形式呈现；它们是情绪类别的输出概率 N。情绪的演变E 彼此依赖，这是由 softmax约束造成的；然而 f*其演变由 s 决定，作为以秒为单位的延迟，它直接受到学习得到的微分斜率的影响—$\frac{dh}{dt}(f^*(t_{-n}))$ 通过 $DOPRI5$ [10] 在前向传播期间求解，进一步作为一般情境下情感感知的一种数学形式。

因此，我们提出一种复杂的情绪智能系统，通过实时交互理解情感轨迹，不仅考察离散模式，还考察由用户在对话交流之间的延迟 s 所决定的模式微分斜率。其结构如下（**右图**）：

$$input_{tok} \rightarrow f^* \rightarrow ICV(f_{\ embedding}\ )_{\frac{dh}{dt}}$$
$$\rightarrow Decoder \rightarrow output_{tok}$$

其中，原始提示工程风格如下（**左图**）：

$$input_{tok} \rightarrow f \rightarrow CraftPrompt\ f(\tilde{E})_{\ t}$$
$$\rightarrow ContextWindow_{\ t} \rightarrow output_{tok}$$

而实际使用的微调解码器本质上为：

$$input_{tok} \rightarrow FineTunedDecoder \rightarrow output_{tok}$$

A. 概念

我们的方法旨在通过时间编码和情感变化追踪，对人类对话中情绪状态的连续流动进行建模，从而实现以心理学为基础的对话建模。

我们提出一种架构，通过对动态演变的特性进行建模，引入隐式情感智能对话中的情绪状态。这项能力支持多种应用，包括数字构造系统（如虚拟代理和情绪响应界面），以及陪伴关系应用（包括照护或治疗场景中的长期人机交互）。该模型强调情感状态的时间性与共同调节，从而为人机对话提供更自然、符合心理学认知的框架。其设计深受认知心理学成熟理论的启发。变色龙效应（*Chartrand* 和 *Bargh*，*1999*）描述了对他人情感表达的无意识模仿，这种现象有助于促进社交互动中的融洽关系与共情 [11]。社会学习理论（*Bandura*，*1977*）在此基础上进一步强调通过观察进行情绪学习，并将自我效能这一概念引入情绪智力发展的过程 [12]。这些理论观点为本系统的情感基础提供了依据，也促使我们追踪对话中的隐性情绪信号。神经科学模型则提供了进一步的结构性启发。情绪感染的神经认知模型（*Prochazkova* 和 *Kret*，*2017*年）指出，情绪共鸣植根于共享的神经激活，尤其是通过感官模仿和共情实现 [13]。这与知觉-行动模型（*Preston* 和 *de Waal*，*2002*）相一致。该模型认为，对他人情绪的感知与相应内部状态的激活之间存在直接映射关系 [14]。研究发现，这些情感反馈机制既存在于面对面沟通中，也存在于媒介化沟通中；近期研究在线互动中相互依存情绪动态的量化（*Luo* 等，*2024*年）进一步为此提供支持，并以定量方式证实了数字平台上的情绪感染现象 [15]。这一发现也建立在早期实证研究的基础之上，包括*Facebook* 情绪感染实验（*2014*年）[16], 。该实验表明，在线环境中情绪基调的细微变化会在社交网络中传播。从计算角度来看，该架构主要受到时间敏感神经建模的启发。液态时间常数网络（*Hasani* 等，*2020*年）通过可学习的衰减常数引入神经元级时间动态（τ），其灵感来自秀丽隐杆线虫的神经生物学，并提供了一种具有生物学依据的机制来编码不断变化的情绪状态 [17]。另一种方法是脉冲神经网络（*Maass*，*1997*），它通过模拟离散的神经元事件来建模符号推理，但其实现通常需要专用的神经形态硬件 [18]。这些时间建模范式支持流畅、具备语境敏感性的对话理解所需的情感调节。通过融合情感心理学与时间动态神经架构，该模型旨在模拟人类对话中的情绪流动性与社会协调。它并非只对静态情绪线索作出反应，而是追踪情感流动随时间发生的变化、转变及其惯性。这使细致的共同调节与语境敏感性成为可能，而这正是情绪智能交互的关键特征；该模型在心理健康技术、自适应对话代理及情感感知领域具有重要应用潜力

机器人学。

II. 连续自然语言处理

预备知识所有数据集预处理均遵循基于深度学习的自然语言处理领域的标准做法及特定领域惯例，相关说明见文献 闲聊 [19]和用于情感识别任务的情感数据集 [20]，不过，为以防万一，我们还进行了额外处理：移除网页链接和单字符话语；最终**16** 行被删除，其中闲聊数据集删除这些行后，我们将这些行对应的对话窗口保留为噪声；而用于情感识别任务的情感数据集最终有**0** 行被删除。

A. 连续数据集

为了有效建模对话交流中情感轨迹的连续性，我们选用了一个开放领域数据集——*chitchat* [19]——该数据集是在一项由大学主办的竞赛中收集的。该数据集包含 $\approx$ 258,145 条话语，每个对话窗口中恰好代表两次真实的学生互动，该窗口定义为 $t_{segment}$ in II-E。经过噪声清理 II后，该数据集已按照时间戳进行时间顺序排序，保留了现实世界对话流程的自然顺序。

B. 情感标注

将每一行记为 t，其中 t_n和 $n \approx 258{,}145$。每一行均使用软标签 $E \in \{e_0, e_1, \ldots, e_N\}$ 进行标注，以表示情感状态。在本研究中，我们将其定义为:

$$E \in \{\text{sadness}_0, \text{joy}_1, \text{love}_2, \text{anger}_3, \text{fear}_4, \text{surprise}_5\}, \quad N = 5$$

接下来，我们定义编码器模型$f(\tilde{x}_t \forall x \in \{\mathcal{A}\})$的一个函数，该模型是针对所有仅包含字符串 $\mathcal{A}$的话语 $\tilde{x}$ 微调后的 ModernBERT [21]。选择 ModernBERT 是因为与前代模型相比，它得益于 RoFormers [22] 中介绍的 RoPE，具有更强的泛化性能，代价是参数略多。因此，设 n表示由 j 列组成的 $\forall e_j \in E \land \ \forall E \in \mathbb{R}^{t_n \times 1}$ 行，并采用混合精度 [23] $float\ \mathcal{F}$;

$$\therefore \underbrace{e_j \in [0,1]_{\mathcal{F}} \forall j \in [0, N] \land E_t \in [0,1]_{\mathcal{F}}}_{softmax(E)} \land f(\tilde{x}_t \forall x \in \{\mathcal{A}\}) \tag{1}$$

$$:= \sum_{t=0}^{t_n} \sum_{j=0}^{e_N} \{e_{tj} | j = 0, \cdots, N\}_t \tag{2}$$

在按照上述定义进行标注之前，模型基于该数据集进行了微调——用于情绪识别任务的情绪数据集 [20]该数据集经过预处理和整理，来源于按照 E 所述方式精确标注的英文 Twitter 消息；训练后的指标恰好 $step = 3200$ 如下:

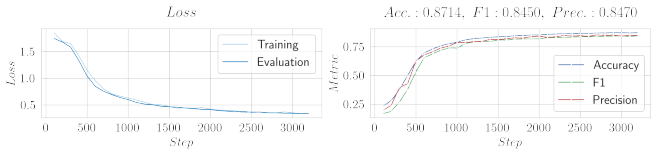

图 2。 $optim:\ Lion, lr:\ 5e-6, lr_{scheduler}:\ cosine,\ batch:$ 32, $\omega\ decay:\ 1e-2\ \forall\theta,\ bf16:$ 1。早停定义为 $\forall t \in [T-5, T] \quad |\mathcal{L}(t) - \mathcal{L}_{\min}| \leq 1e-2$，其中 5 表示步数 $\neq$ 轮。

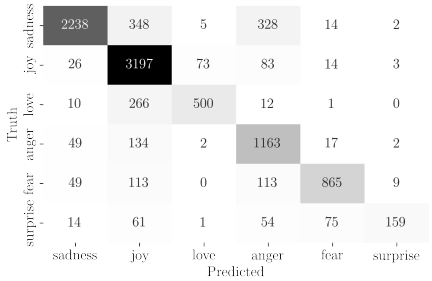

图 3. $surprise$ 标签表示预测数量最少，这源于数据集中的类别不平衡 [20]，不过，我们通过 $Qwen2.5$ [24] 对每个标签 $\{love, anger, fear, surprise\}$进行生成式释义增强来缓解这一问题，使样本数量增加了一倍，但类别仍然不平衡。

C. 循环变换

将已标注的——*chitchat* 数据集中格式为 yyyy:MM:dd:hh:mm:ss 的原始时间戳进行周期性转换，对于时间建模至关重要。为避免时间格式产生歧义，只取 hh:mm:ss，随后我们可以定义一个三角函数，其中$\mathcal{R}$ 表示弧度，记作 $\forall \mathcal{R} \in \mathbb{R}^{t_n \times j} \land \forall j \in \{sin, cos\}$，因此每一行 $\mathcal{R}$定义如下;

$$\therefore \mathcal{R} \in \{sin, cos\}_t \tag{3}$$

$$:= \sum_{t=0}^{t_n} \sum_{j=0}^{j=1} ((\frac{2\pi \cdot hh_t}{24})_j, (\frac{2\pi \cdot mm_t}{60})_j, (\frac{2\pi \cdot ss_t}{60})_j)_t \tag{4}$$

因此，我们将延迟列 τ 记为 $\tau \in \{ss_t - ss_{t-1}\} \forall t \neq 0 \land \ \forall \tau \in \mathbb{R}^{t_n \times 1}$。具体而言， ss 在计算 τ 之前不 应进行循环变换，而应在计算 τ之后**再** 进行循环变换，因为 τ 应采用 ss 的绝对形式，以便从 $[t_0, t_n]$ $T =$连续构造，$\int_{t_0}^{t_n} f\ t\ dt$ 从而实现 ( ) 的真正单调连续积分。因此，我们定义一个离散积分的累加表达式;

$$\therefore \tau_{t_0} : [0] \land \forall \tau_{t \neq 0} \in [1.00, 86400.00] \land \tilde{T} : \int_{t_0}^{t_n} f(t) dt \tag{5}$$

$$:= \sum_{t=0}^{t_n} \tau_t |ss_t - ss_{t-1}| \tag{6}$$

D. 情感幅度

情感幅度 Δ 指的是情绪变化 这一统称，通过计算在 t 和 $t-1$ 处情感状态的差异来表示，记为 $\forall\Delta \in \mathbb{R}^{t_n\times 1}$；具体而言，只需计算 L1范数，并将每一行定义为：

$$\therefore e_j\Delta_{t_0} \in [0] \wedge \forall e_{tj} \in [0,1], \Delta_t \tag{7}$$

$$:= \sum_{t=1}^{t_n}\sum_{j=0}^{e_N} |e_{tj} - e_{tj-1}| \tag{8}$$

E. 连续插值

将离散数据建模为连续轨迹，需要使用 $\tilde{T} = \int_{t_0}^{t_n} f(t)dt$ 方程II-C 中的 **`scipy.interpolate.CubicHermiteSpline`** [25],˜，需要注意的是 T 是表示为 $\tilde{T} \in \mathbb{C}^{\tilde{t}_n\times 1}$的合成全局变量，而 非 *Hermite* 三次样条插值会因 $S_i(t)$ 的 C^2 平滑性而产生不必要的离群值，其过冲过渡斜率会˜影响 T，形式化表示为：

$$\tilde{T}_{C^{1\vee 2}} \approx \sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} S_i(t)dt \tag{9}$$

$$:= [f(t_i)t_{i+1} - t_i \tag{10}$$

$$+\frac{f(b_i)t_{i+1}-t_i^2}{2} + \frac{f(c_i)t_{i+1}-t_i^3}{3} + \frac{f(d_i)t_{i+1}-t_i^4}{4}] \tag{11}$$

$$\underbrace{f(c_i)}_{2nd\ derivative} = (t_{i+1}-t_i)_{i-1}(c_{i-1} + 2((t_{i+1}-t_i)_{i-1} \tag{12}$$

$$+(t_{i+1}-t_i))c_i + (t_{i+1}-t_i)c_{i+1} \tag{13}$$

$$= 3(\frac{f(t_{i+1})-f(t_i)}{t_{i+1}-t_i} - \frac{f(t_i)-f(t_{i-1})}{(t_{i+1}-t_i)_{i-1}}, \tag{14}$$

$$\underbrace{f(b_i)}_{1st\ derivative} = \frac{f(t_{i+1})-f(t_i)}{t_{i+1}-t_i} - \frac{t_{i+1}-t_i}{3}(2c_i + c_{i+1}) \tag{15}$$

$$\underbrace{f(d_i)}_{coefficient} = \frac{c_{i+1}-c_i}{3(t_{i+1}-t_i)}, \quad \therefore v_i = f(t_i) \tag{16}$$

因此，在我们的 *chitchat* 案例中，引入时间分段至关重要，即使对于三次 Hermite 样条也是如此；通过将斜率$\forall t_{C^1} \in [t_{start}, t_{end}]$ 和 C^0 封装在 $t_{segment}$之间，可以防止全局异常值产生的噪声。分段插值会计算出独特的微分斜率模式序列，由于局部效应，该序列有别于其他分段，同时保留其连续性。因此，我们将每个分段定义为具有隐式相互依赖变量的形式；

$$v = t_{segment} \in \{E, \mathcal{R}_{sin,cos}, \tau, \Delta\}_t \tag{17}$$

$$\wedge\forall\ t_{segment} \in [t_{start}, t_{end}] \tag{18}$$

$$,\quad \text{where} \quad E \in \{e_0, \cdots, e_N\} \tag{19}$$

$$,\quad \mathcal{R} \in \{\tilde{h}, \tilde{m}, \tilde{s}\}_{sin,cos} \tag{20}$$

因此，我们可以通过对 S_0处的前向差分、 S_i处的中心差分以及 S_{n-1} 处的后向差分进行插值，来计算斜率 S ，其定义如下；

$$forward\ _{S_0} = \frac{v_1 - v_0}{t_1 - t_0} \tag{21}$$

$$,central\ _S = \frac{v_{i+1} - v_{i-1}}{t_{i+1} - t_{i-1}} \tag{22}$$

$$,backward\ _{S_{n-1}} = \frac{v_n - v_{n-1}}{t_n - t_{n-1}} \tag{23}$$

$$\text{finally,} \quad \forall\ t_{segment} \in \{t_i, t_{i+1}\} \equiv \tilde{T}_{C^{0\vee 1}} \in \{\tau\} \tag{24}$$

$$\int_{t_0}^{t_n} f(t) = h_{00}(x)\cdot v_i \text{ where } \quad h_{00}(x) = 2x^3 - 3x^2 + 1, \tag{25}$$

$$+h_{10}(x)\cdot(t_{i+1}-t_i)\cdot m_i \text{ where } \quad h_{10}(x) = x^3 - 2x^2 + x, \tag{26}$$

$$+h_{01}(x)\cdot v_{i+1} \text{ where } \quad h_{01}(x) = -2x^3 + 3x^2, \tag{27}$$

$$+h_{11}(x)\cdot(t_{i+1}-t_i)\cdot m_{i+1} \text{ where } \quad h_{11}(x) = x^3 - x^2 \tag{28}$$

$$, \quad \text{and} \quad x = \frac{t - t_i}{t_{i+1} - t_i} \tag{29}$$

F. CEmoFlow数据集

我们介绍了 CEmoFlow数据集，其结构由按时间顺序排列的对话交流构成，其中的话语已完全匿名化，并去除了*chitchat* [19]。该数据集提供连续插值的情绪软标签及其时间戳，有助于使用物理信息神经网络（PINN）模拟现实世界中的情绪动态。

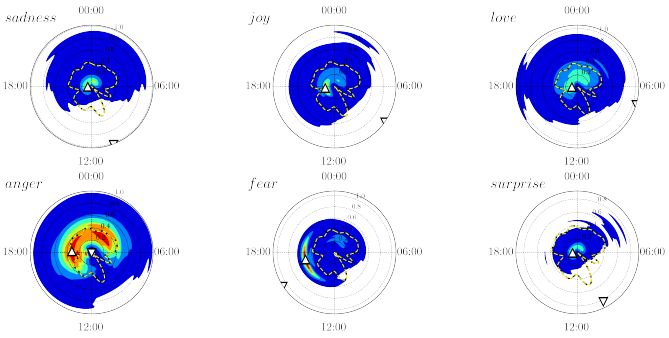

图 4. 高斯核密度估计，包含 250.000 个样本，其中 $hh_{sin,cos}$ 转换为 $arctan2(hh_{sin,cos})$，表示 $x \in${上午 中午,下午 中午} 轴上每个软标签分布；沿着 $y \in \forall e_N \in [0,1]$ 轴上的等高线，依据其强度峰值绘制。黄色线表示 $\mu\Delta \in [0,1]$（归一化），而 $\triangle$ 标记分布最高值， $\triangledown$ 标记等高线上的最低分布密度。

X	$count$	$mean$	std	min	25%	50%	75%	$max\tilde{r}$
	6.76e+06	3.16e+06	1.89e+06	0.00e+00	1.64e+06	2.99e+06	4.78e+06	6.61e+06
	6.76e+06	4.37e+01	6.62e+01	-5.35e+01	6.93e+00	1.94e+01	5.00e+01	5.59e+02
e_0	6.76	+06	8.82	02	8.71	02	1.18	01
	3.97	02	6.77	02	1.08	01	1.17	+00
e_1	6.76	+06	1.57	01	1.24	01	1.25	01
	7.51	02	1.26	01	1.99	01	1.09	+00
e_2	6.76	+06	1.77	01	1.36	01	1.22	01
	8.59	02	1.37	01	2.28	01	1.10	+00
e_3	6.76	+06	2.89	01	1.47	01	9.60	02
	1.84	01	2.70	01	3.69	01	1.14	+00
e_4	6.76	+06	2.06	01	1.36	01	1.20	01
	1.04	01	1.81	01	2.81	01	1.13	+00
e_5	6.76	+06	8.30	02	6.96	02	1.16	01
	4.30	02	6.96	02	1.02	01	9.84	01
Δ	6.76	+06	6.78	01	3.35	01	2.50	01
	4.43	01	6.37	01	8.77	01	2.19	+00
$\mathcal{R}\ sin\ h$	6.76	+06	1.34	01	7.02	01	1.29	+00
	8.66	01	2.59	01	5.00	01	1.29	+00
$\mathcal{R}\ cos\ h$	6.76	+06	4.23	01	5.52	01	1.30	+00
	0.00	+00	5.21	01	8.66	01	1.29	+00
$sin\ m$	6.76	+06	1.86	02	6.95	01	1.31	+00
	6.98	01	9.90	03	6.69	01	1.33	+00
$cos\ m$	6.76	+06	2.13	02	7.09	01	1.27	+00
	6.76	01	6.99	02	7.21	01	1.30	+00
$sin\ s$	6.76	+06	1.03	02	6.44	01	1.40	+00
	5.96	01	1.58	02	5.77	01	1.39	+00
$cos\ s$	6.76	+06	6.61	03	6.38	01	1.38	+00
	5.72	01	8.10	03	5.86	01	1.41	+00

III. 局限性

CEmoFlow 以及采用物理信息神经网络为文本生成大语言模型的潜在空间引导 [6]所带来的局限性显著增加了实现复杂度。实验采用非循环多输出回归神经常微分方程 [8]，并使用 $y \in \{e_0, e_1, e_2, e_3, e_4, e_5\}$进行训练，其中 $X \subset y$ 导致计算时间显著增加，使用$Dopri5$训练 `121:45:12` 直至通过早停达到收敛；相比之下，在相同训练配置下，多输出回归长短期记忆网络仅耗时 `9:11:56` 。该模型唯一的局限性是其黑箱特性。实验使用`RTX4070` ，其包含4608 个 CUDA核心和144 个张量核心。不过，ODE 的求解完全由中央处理器 完成，之后再由图形处理器塑造潜在空间。

参考文献

[1]L. Christ、S. Amiriparian、M. Milling、I. Aslan 和 B. W. Schuller，《利用变换器和弱监督学习对书面故事中的情绪轨迹进行建模》，2024年。 [在线]。可访问：https://arxiv.org/abs/ 2406 02251 。A. [2] Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、L. Kaiser 和 I. Polosukhin，《注意力就是你所需要的一切》，第 5998–6008 页，2017年–。[3] Y. Li，《关于用于上下文学习的零样本提示设计的实用综述》，载于《自然语言处理最新进展会议——自然语言处理大语言模型》，RANLP 系列。INCOMA Ltd.，舒门，保加利亚，2023年，第 641–647 页。 [在线]。可访问：http://dx.doi.org/ 10 26615 978 954 452 092 2 069 。/ [4] - - - - H. Ma 、 C. Zhang 、 Y. Bian 、 L. Liu 、 Z. Zhang 、 P. Zhao 、 S. Zhang、H. Fu、Q. Hu 和 B. Wu，《面向大语言模型的公平性引导的少样本提示》，2023年。 [在线]。可访问：https: //arxiv.org/abs/ 2303 13217 。B. Conway-Smith 和 R. L. West，《抽离式正 [5] 念的计算机制》，2024年。 [在线]。可访问：https://arxiv.org/abs/ 2409.15289[6] S. Liu、H. Ye、L. Xing 和 J. Zou，《上下文向量：通过潜在空间引导使上下文学习更有效、更可控》， 2024 [ ] 2311 06668 。在线。 [7] 可访问：https://arxiv.org/abs/ 。K. Funahashi 和 Y. Nakamura，《通过连续时间循环神经网络逼近动力系统》，《神经网络》，第 6 卷，第 6 期，第 801–806 页，1993年。[8] R. T. Q. Chen、Y. Rubanova、J. Bettencourt 和 D. Duvenaud，《神经常微分方程》，《神经信息处理系统进展》，2018年。[9] P. Kidger、J. Morrill、J. Foster 和 T. Lyons，《用于不规则时间序列的神经控制微分方程》，载于《神经信息处理系统进展》，第 33 卷，2020年。

[10]J. Dormand 和 P. Prince，《一族嵌入式 Runge-Kutta 公式》，《计算与应用数学杂志》，第 6 卷，第 1 期，第 19–26 页，1980年。 [在线]。可访问：https: // www.sciencedirect.com/science/article/pii/ 0771050 80900133 X J. L. Lakin、V. [11] E. Jefferis、C. M. Cheng 和 T. L. Chartrand，《变色龙效应作为社会黏合剂：无意识模仿的进化意义证据》，《非言语行为杂志》，第 27 卷，第 3 期，第 145–162 页，2003年。[12] G. F. Jensen，《犯罪行为社会学》，载于《国际社会与行为科学百科全书》（第二版），第二版，J. D. Wright 编。牛津：爱思唯尔，2015年，第 68–74 页。[13] E. Prochazkova 和 M. E. Kret，《通过模仿连接心智并分享情绪：情绪感染的神经认知模型》，《神经科学与生物行为评论》，第 80， 99 114, 页–。在线。可访 2017 [ ] 问：https://www.sciencedirect.com/science/article/ pii/S 0149763416306784 [14] S. D. Preston 和 F. B. De Waal，《共情：其终极与近因基础》，《行为与脑科学》，第 25 卷，第 1 期，第 1–20 页，2002年。[15] Y. Luo、D. Sornette 和 S. C. Lera，《在线互动中自激情绪动力学的量化》，2024年。 [在线]。可访问：https://arxiv.org/abs/ 2408 05700 。A. D. I. [16] Kramer、J. E. Guillory 和 J. T. Hancock，《通过社交网络实现大规模情绪感染的实验证据》，《美国国家科学院院刊》，第 111 卷，第 24 期，第 8788–8790 页，2014年。 [在线]。可访问：https://www.pnas.org/doi/abs/ 10 1073 1320040111 。/pnas。R. Hasani、M. [17] Lechner、A. Amini、D. Rus 和 R. Grosu，《液态时间常数网络》，《AAAI 人工智能会议论文集》，第 35 卷，第 9 期，第 7657–7666 页，2021年 5 月。 [在线]。可访问：https://ojs.aaai.org/index.php/AAAI/article/view/ 16936 [18] W. Maass，《脉冲神经元网络：神经网络模型的第三代》，《神经网络》，第 10， 9， 1659 1671, 卷，第 1997 [ ] 10， 9， 1659 1671, 期，第–页。在线。 1997 [ ] 可访问：https://www.sciencedirect.com/science/article/ pii/S 0893608097000117 [19] W. Myers、T. Etchart 和 N. Fulda，《对话支架：开放域对话中响应优先级排序的基于类比的方法》，《第 12 届国际智能体与人工智能会议论文集》，2020年。 [在线]。可访问：https://www.scitepress.org/Papers/2020/89399/89399.pdf[20] E. Saravia、H.-C. T. Liu、Y.-H. Huang、J. Wu 和 Y.-S. Chen，《CARER：用于情绪识别的情境化情感表征》，载于《2018年自然语言处理实证方法会议论文集》，E. Riloff、D. Chiang、J. Hockenmaier 和 J. Tsujii 编。比利时布鲁塞尔：计算语言学协会，2018年 10–11 月，第 3687–3697 页。 [在线]。可访问：https://aclanthology.org/D18-1404/ [21] B. Warner、A. Chaffin、B. Clavi´e、O. Weller、O. Hallstr¨om、S. Taghadouini、A. Gallagher、R. Biswas、F. Ladhak、T. Aarsen、N. Cooper、G. Adams、J. Howard 和 I. Poli，《更智能、更优秀、更快速、更持久：用于快速、内存高效和长上下文微调与推理的现代双向编码器》，2024年。 [在线]。可访问：https://arxiv.org/abs/ 2412 13663 。J. Su、Y. Lu、S. Pan、 A. Murtadha、B. Wen 和 Y. Liu，《Roformer：采用旋转位置嵌入的增强型 [22] Transformer》，2023年。 [在线]。可访问：https://arxiv.org/abs/ 2104 09864 。P. Micikevicius、S. Narang、J. [23] Alben、G. Diamos、E. Elsen、D. Garcia、B. Ginsburg、M. Houston、O. Kuchaiev、G. Venkatesh 和 H. Wu，《混合精度训练》，载于《国际学习表征会议》，2018年。 [在线]。可访问：https: //openreview.net/forum?id=r1gs9JgRZ [24] 通义千问，：，A. Yang、B. Yang、B. Zhang、B. Hui、B. Zheng、B. Yu、C. Li、D. Liu、F. Huang、H. Wei、H. Lin、J. Yang、J. Tu、J. Zhang、J. Yang、J. Yang、J. Zhou、J. Lin、K. Dang、K. Lu、K. Bao、K. Yang、L. Yu、M. Li、M. Xue、P. Zhang、Q. Zhu、R. Men、R. Lin、T. Li、T. Tang、T. Xia、X. Ren、X. Ren、Y. Fan、Y. Su、Y. Zhang、Y. Wan、Y. Liu、Z. Cui、Z. Zhang 和 Z. Qiu，《Qwen2.5 技术报告》，2025年。 [在线]。可访问：https://arxiv.org/abs/ 2412 15115 。P. Virtanen、R. Gommers、T. E. [25] Oliphant、M. Haberland、T. Reddy、D. Cournapeau、E. Burovski、P. Peterson、W. Weckesser、J. Bright、S. J. van der Walt、M. Brett、J. Wilson、K. J. Millman、N. Mayorov、A. R. J. Nelson、E. Jones、R. Kern、E. Larson、C. J. Carey、 ˙I. Polat、Y. Feng、E. W. Moore、J. VanderPlas、D. Laxalde、J. Perktold、R. Cimrman、I. Henriksen、E. A. Quintero、C. R. Harris、A. M. Archibald、A. H. Ribeiro、F. Pedregosa、P. van Mulbregt 和 SciPy 1.0 贡献者，《SciPy 1.0：基础

《Python 科学计算算法》，*Nature Methods*，第 17 卷，第 261–272 页，2020年。