arXiv:2602.06394v1 [cs.AI] 6 Feb 2026

通过质量感知分词解锁用于基础模型预训练的含噪真实世界语料库

Arvid E. Gollwitzer [1 2 3] **Paridhi Latawa** [2] **David de Gruijl** [3] **Deepak A. Subramanian** [1 4] **Adrián Noriega de la Colina** [1 2 5 6]

摘要

当前的分词方法处理序列数据时未考虑信号质量，因此在含噪真实世界语料库上的效果有限。我们提出*QA-Token*（质量感知分词），将数据可靠性直接纳入词汇表构建。我们有三项主要贡献：(i) 一种双层优化形式，联合优化词汇表构建与下游性能；(ii) 一种强化学习方法，通过质量感知奖励学习合并策略，并提供收敛保证；(iii) 一种通过 Gumbel-Softmax 松弛实现端到端优化的自适应参数学习机制。实验评估表明，该方法持续带来提升：基因组学（变异检测的 F1 分数比 BPE 高 6.7 个百分点），金融（夏普比率提升 30%）。在基础模型规模下，对 METAGENE-1 的 1.7 万亿碱基对语料库重新分词，实现了最先进的病原体检测效果（马修斯相关系数为 94.53），同时将词元数量减少 15%。我们让涵盖拍碱基级基因组序列和数 TB 金融时间序列的含噪真实世界语料库可用于基础模型训练，且不会增加任何推理开销。

1. 引言

分词是原始数据与神经计算之间的接口。当前的方法，例如字节对编码 (BPE)(Sennrich 等人，2016年)，完全依赖频率统计，并假设出现频率与语义重要性相关。当数据质量差异显著时，这一假设便不再成立——从基因组学中的测序错误 (Ewinget al., 1998)到金融市场中的微观结构噪声 (Andersenetal., 2001年)。使用基于频率的分词在嘈杂语料库上训练的模型会继承这些错误，导致性能下降——这一效应如今已通过质量感知缩放定律得到形式化描述 (Subra-manyam et al.，2025)。

问题十分严峻：第三代测序的错误率超过 10%（Wenger 等人，2019），然而当前的分词器对高置信度区域和易出错区域一视同仁。在金融领域，超过 40% 的高频数据含有微观结构噪声（Hansen & Lunde，2006），但分词方法无法区分信号质量。这一局限制约了利用真实世界数据进行基础模型训练。

可用生物数据的规模放大了这一挑战。公共序列存储库目前包含超过67拍字节对（Pbp）的原始测序数据，欧洲核苷酸档案库大约每45个月翻一番（Karasikov 等人，2025）。高效索引的最新进展已使这些拍字节级档案库能够以低至每查询兆碱基对0.74美元的成本进行全文检索，表明大规模序列分析的基础设施正在迅速成熟。然而，由于质量异质性，这些数据中相当大的一部分仍未充分用于基础模型训练——标准的基于频率的分词方法要么完全丢弃低质量读段，要么将测序错误传播到学习得到的表征中。数据可得性与可用性之间的这一差距，促使人们从根本上重新思考分词如何处理质量变化。

我们提出**质量感知分词（QA-Token）**，这是一个将数据质量纳入词汇表构建的框架。我们做出三项关键贡献：

1. 复杂度分析下的双层优化：我们将分词形式化为一个双层优化问题（定理 3.1），该问题联合优化词汇表构建与下游性能。我们证明该问题是NP难问题（定理 3.2），并开发了一种有理论保证的原理性近似方

[1]麻省理工学院与哈佛大学布罗德研究所，美国马萨诸塞州剑桥 [2] 麻省理工学院，美国马萨诸塞州剑桥 [3]Anto Biosciences（YC F25）[4]麻省理工学院科赫综合癌症研究所，美国马萨诸塞州剑桥 [5] 麦吉尔大学神经病学与神经外科系，加拿大蒙特利尔 [6]蒙特利尔神经病学医院研究所，加拿大蒙特利尔。通信作者：Arvid E. Gollwitzer <arvidg@mit.edu>。

预印本。*2026*年*2*月*9*日。

具有理论保证的mation方案。

2. 具备收敛保证的强化学习：我们将词汇表构建建模为马尔可夫决策过程（定理 E.4），并采用强化学习来发现最优合并策略。我们给出了形式化的收敛分析（定理 E.5），并实现了与最优自适应策略相差 ($1-1/e$) 的近似。

3. 可微参数学习：通过 Gumbel-Softmax 松弛（定理 C.8），我们实现了质量敏感度参数的端到端学习，并证明了其一致性和梯度有界性（定理C.7）。

我们证明，QA-Token 在噪声条件下达到了信息论最优性（定理 C.13），为质量感知分词提供了形式化依据。评估结果显示，在算法交易中的夏普比率提高了 30%，基因组变异位点检测 F1 分数提高了 6.7 个百分点（BPE 为 0.824，而本方法为 0.891）；集成到 70 亿参数基础模型中时，也实现了最先进性能。

核心贡献： (i) 我们推导出一种质量感知的合并分数（定理C.3），通过可学习敏感度平衡频率、质量与领域约束 α （第 C.2节）。(ii) 我们将词汇表构建形式化为一个马尔可夫决策过程（定理 E.4，第 E.7节），并通过自适应次模性实现 $(1-1/e)$-近似。(iii) Gumbel-Softmax 松弛支持端到端参数学习，且具有 $O(1/\sqrt{T})$ 收敛速率（定理 E.2，第 E.4节）。(iv) 领域特定实例化在 15+ 个基准测试上达到最先进性能。

我们的分析表明，将质量信号纳入分词后，即使在基于频率的方法失效的嘈杂语料库上也能进行训练，从而扩大可用于训练基础模型的数据范围，并带来更广泛的科学和经济影响（第7.1节）。

2. 噪声领域的质量指标

质量指标必须满足三个形式化属性，才能以有原则的方式整合到合并分数中：(i) 有界性 ($q \in [0,1]$)，以确保数值稳定性；(ii) *Lipschitz* 连续性，以便在自适应学习期间稳定地计算梯度；(iii)噪声注入下的单调性 （噪声越高，质量越低），以确保语义一致性。我们证明，这些属性对于我们的领域特定实例化均成立（定理C.1，第C.1节）。

基因组学： 我们采用带位置调整衰减的 Phred 分数：$q'_{s_j} = q_{s_j} \cdot \exp(-\beta_{\text{pos}} \cdot j/L)$，其中 β_{pos} 是学习得到的，L 是读长。词元质量通过几何平均值 $q_t = (\prod_{j=1}^{|t|} q'_{s_j})^{1/|t|}$ 进行聚合，以确保对低质量区域保持敏感——单个不可靠碱基就会损害整个词元（公式 13，第D节）。

金融： 我们结合市场微观结构的四个维度：(i) 流动性 q_{liq} （买卖价差、深度），(ii) 信号质量 q_{sig} （价格变动的信噪比），(iii) 稳定性 q_{stb} （波动率市场状态），以及 (iv) 信息含量 q_{info} （订单流的信息性）。综合分数 $q_t^{\text{finance}} = \sum_k w_k q_{k,\,t}$使用学习得到的权重；算术平均聚合反映了金融数据的加性噪声特征（公式14，第 D节）。

定义好质量指标后，我们现在将其如何整合到分词目标函数中形式化。

3. QA-Token 的数学形式化

3.1. 符号与设定

令 $\mathcal{S} = \{S_1, S_2, \ldots, S_N\}$ 表示一个包含 N 个序列的语料库，其中每个序列 $S_k = (s_k,1, \ldots, s_k,n_k)$ 由基础字母表 Σ 中的元素构成。每个原子元素 s_k,i 都关联一个归一化质量分数 $q_k,i \in [0,1]$，其定义见第 2 节。初始词汇表定义为 $V_0 = \Sigma$。在分词过程的任意步骤 k， V_k 表示当前词汇表。对于任意词元 $a \in V_k$，我们将其在语料库中的频率记为 $f(a)$；对于一个相邻对 (a,b)，其共现频率记为 $f(a,b)$。词元 t 的原子单位长度为 $|t|$。令 q_t 表示词元 t 的聚合标量质量，该值通过特定领域聚合函数计算得出（见第 D 节）。

3.2. 正式的问题定义与目标函数

我们将分词形式化为寻找一个分词器 $\mathcal{T}$，使其最大化目标函数 $\mathcal{J}$，同时兼顾下游任务表现、词汇复杂度和数据可靠性。令$\mathcal{S} = \{S_1, S_2, \ldots, S_N\}$ 表示从底层数据分布 $\mathcal{P}_{\text{data}}$中采样得到的、包含 N 个序列的语料库，其中每个 $S_k = (s_{k,\,1}, \ldots\ldots, s_{k,\,n_k})$均由基础字母表 Σ 中的元素构成。分词器 $\mathcal{T}: \mathcal{S} \to \mathcal{Z}$ 使用词汇表 V将语料库映射为切分结果 $\mathcal{Z} = \{Z_1, \ldots, Z_N\}$。

定义3.1 （双层分词问题）。最优的质量感知分词问题可表述为以下双层优化问题：

$$\max_{\mathcal{T} \in \mathcal{G}(K)} \mathcal{J}(\mathcal{T}) := \lambda_{\text{LM}}\, \mathcal{L}_{\text{LM}}(\mathcal{T}) - \lambda_{\text{comp}}\, \Phi(V) + \lambda_{\text{qual}}\, Q(V, \mathcal{Z}), \tag{1}$$

其中，语言模型的表现为：

$$\mathcal{L}_{\text{LM}}(\mathcal{T}) = \max_{\theta \in \Theta} \mathbb{E}_{\mathcal{D} \sim \mathcal{P}_{\text{data}}}[\log p_\theta(\mathcal{D}|\mathcal{T})], \tag{2}$$

以及 $\mathcal{G}(K) = \{\mathcal{T}: \ |V_\mathcal{T}| - |\Sigma| \le K\}$ 表示从基础字母表 Σ 出发，通过至多 K 次合并操作可达的分词器集合，其中 Θ 是语言模型的参数空间。

目标函数 $\mathcal{J}$ 平衡了三个组成部分：(i) 下游性能 $\mathcal{L}_{\text{LM}}(\mathcal{T})$，即最大化预期对数似然；(ii) 复杂度惩罚 $\Phi(V) = |V| \log |V| + \sum_{t \in V} |t| \cdot H(t)$，遵循最小描述长度原则（Rissanen，1978）——第一项惩罚词汇表大小（词元索引的描述长度），第二项则通过经验熵 $H(t) = -\sum_{\sigma \in \Sigma} \frac{n_\sigma(t)}{|t|} \log \frac{n_\sigma(t)}{|t|}$ 惩罚词元内部的复杂度，该经验熵衡量词元 t 中原子元素的复杂度（其中 $n_\sigma(t)$表示元素 σ 的计数； $H(t) = 0$ 适用于单元素词元）；以及 (iii) 可靠性奖励 $Q(V, \mathcal{Z}) = \frac{1}{\sum_{k=1}^{N} |Z_k|} \sum_{k=1}^{N} \sum_{t \in Z_k} g(q_t)$，通过凹函数 g汇总词元质量。

聚合函数 g 呈现凹性，以捕捉合并高质量组成部分时的边际收益递减。在本文中，我们采用 $g(x) = (x + \epsilon_Q)^\alpha$，其中$0 < \alpha < 1$（严格凹），并采用 $\epsilon_Q = 10^{-8}$ 以确保数值稳定性。边界情形 $\alpha = 1$ 可实现线性聚合，这适用于质量贡献具有可加性、而非受到边际收益递减影响的情况。

定理 3.2（计算复杂度）。式 *1* 中的双层优化问题一般而言是 *NP*难的（*Dempe，2020*）；事实上，多项式双层规划是 Σ_2^p 难的（*Cen & Chi，2023*年），其在多项式层次中比 *NP* 高一层。最坏情况下需要$O(|\Sigma|^K \cdot K! \cdot N \cdot n \cdot |\Theta|)$ 次计算（证明见第 *C.5*节）。

鉴于这种计算难处理性，我们提出一种有原则的近似方案，将贪心合并选择与强化学习相结合，具体内容将在后续章节中详述。

3.3. 质量感知合并评分

我们通过以下结果将质量信号纳入基于点互信息的词元化：

定理 3.3(质量感知合并评分)。 使目标函数增量 $\Delta\mathcal{J}$ 的一阶近似最大化的最优贪心合并分数为：

$$w_{ab} = \frac{f(a,b)}{f(a)f(b) + \epsilon_f} \cdot (\bar{q}_{ab} + \epsilon_Q)^\alpha \cdot \psi(a,b) \qquad (3)$$

其中 $\bar{q}_{ab} = (q_a + q_b)/2$ 对组成部分的质量取平均值，$\alpha \in (0, 1]$ 控制质量敏感度，而 $\psi(a,b)$ 表示领域约束。（证明采用一阶近似，见第 *C.2*节。）

该分数综合了统计关联（点互信息项）、数据可靠性（质量项）和领域特定要求。有界性和利普希茨连续性的证明见定理 C.4 （第 C.5 节）。

4. 学习框架：强化学习与自适应参数

我们将词汇表构建视为一个分两个连续阶段的学习问题。**阶段 1** （强化学习策略优化）使用带有质量感知奖励的近端策略优化（PPO）学习合并选择策略 π_{θ_π}，并保持初始参数 $\theta^{(0)}$ 为固定值。**阶段 2** （自适应参数学习）通过 Gumbel-Softmax 松弛优化θ_{adapt}，以提升下游性能；该阶段使用贪心模拟和复合逻辑值ℓ_{ab}（θ_{adapt}），而非直接调用强化学习策略——阶段 1 的策略用于初始化候选合并，并提供方差缩减基线。梯度 $\nabla_{\theta_{\text{adapt}}} L_{\text{task}}$通过 Gumbel-Softmax 合并选择进行传播，从而实现端到端学习（详见第 E 节，算法 1–3）。

4.1. 强化学习形式化

定义 4.1 （分词化马尔可夫决策过程）。词汇表构建马尔可夫决策过程为 $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, T)$，其中：状态$s_t \in \mathcal{S}$ 编码当前词汇表、合并候选项和语料库统计信息；动作 $a_t \in \mathcal{A}_t$ 选择合并对；转移 $\mathcal{P}$ 是确定性的词汇表更新；奖励 $\mathcal{R}$具有质量感知特性（第 4.2 节）；$\gamma \in (0, 1]$ 是折扣因子； T是时域（目标词汇表大小减去基础字母表大小）。完整定义见第 E.7 节。

强化学习目标是寻找策略 π_{θ_π}： $\mathcal{S} \to \Delta(\mathcal{A})$，使其在 T 次操作中使用近端策略优化（Schulman 等人，2017年）最大化期望累积奖励，并依据（Bhandari 和 Russo，2021年；Cen 和 Chi，2023年）获得全局收敛保证。定理 E.5 （第 E.7 节）证明了马尔可夫决策过程良构性。

4.2. 奖励函数设计

多目标奖励 $R(a, b;\, \theta^{(0)}\,\text{adapt}) = \sum_j \lambda_j \hat{} R_j(a,b)$ 综合了质量、信息、复杂度和领域特定组件。每个原始奖励 R_j^{raw} 都通过自适应运行统计量和指数移动平均进行归一化：$\mu_{j,t}^{\text{run}} = (1 - \beta_{\text{norm}})\mu_{j,t-1}^{\text{run}} + \beta_{\text{norm}} R_j^{\text{raw}}$，从而得到 $\hat{} R_j = (R_j^{\text{raw}} - \mu_{j,t-1}^{\text{run}})/(\sigma_{j,t-1}^{\text{run}} + \epsilon_R)$。这确保了在非平稳策略优化过程中，奖励有界且尺度不变（定理 C.5，E.8节）。

4.3. 分词参数的自适应学习

经过强化学习优化后，我们通过 Gumbel-Softmax（Jang 等人，2017；Maddison等人，2017）学习 θ_{adapt} （质量敏感度 α、领域因子 $\beta_{\text{pos}}/\beta_{\text{vol}}$、权重），以最小化$L_{\text{total}}$（$\theta_{\text{adapt}}$） $= L_{\text{task}}$（θ_{adapt}） $+ \lambda_{\text{reg}} \|\theta_{\text{adapt}}\|_2^2$。温度退火 $\tau(t)$ $= \tau_{\text{init}} \exp(-\beta_{\text{anneal}} t / T_{\text{anneal}})$可确保收敛（定理 C.7、定理 E.2；E.9节、E节）。两阶段框架——先以固定 $\theta^{(0)}$ 进行强化学习优化，再进行自适应学习——最终通过 w_{ab}（a、 b; θ^*_{adapt}）贪心地构建词汇表（E节，算法 1–

3).

4.4. 双时间尺度收敛

序列优化遵循双时间尺度随机逼近：策略在快速时间尺度上更新（学习率 $\eta_\pi^{(t)}$），自适应参数在慢速时间尺度上更新（$\eta^{(t)}$ 自适应），其中 $\eta_\pi^{(t)}/\eta^{(t)}$ 自适应 $\to \infty$ 作为 $t \to \infty$。在假设 (A1)–(A4) 下，该过程收敛到一个局部纳什均衡，其中 θ_π^* 最大化 $J(\pi;\ \theta_{\text{adapt}}^*)$ 而 θ_{adapt}^* 最小化 $L_{\text{total}}(\theta_{\text{adapt}};\ \pi^*)$。趋近全局最优解的质量界和初始化策略详见 E节。

4.5. 理论保证

我们的框架在假设（A1）–（A4）下提供以下保证，详见第C.6节：（i）有界/Lipschitz 合并分数 w_{ab}（定理C.4），（ii）具有严格为正的运行标准差的稳定指数移动平均归一化（定理C.5），（iii）近端策略优化收敛到驻点（定理C.6），（iv）一致且有界的 Gumbel-Softmax 梯度（定理C.7），以及（v）（$1-1/e$）-对最优自适应策略的近似（通过自适应次模性）。

信息论最优性：基于信息瓶颈理论（Tishby 等人，1999；Alemi 等人，2017），我们的分析（定理C.13，C.9节）表明 QA-Token 最小化质量感知信息瓶颈：$\mathcal{L}_{\text{QA}}(V) = -I(T;\ Y|Q) + \beta \cdot I(T;\ X|Q)$，并在以质量 Q 为条件的情况下，通过最大化任务相关信息 $I(T;\ Y|Q)$，同时最小化冗余表示复杂度 $I(T;\ X|Q)$，在噪声条件下实现最优压缩-相关性权衡。完整证明见 C.5 节和 E节。

在建立理论框架并给出收敛保证后，我们在具有不同噪声特征的两个领域中对 QA-Token 进行了实证验证。

5. 实证验证

设置： 结果为 10 次试验的均值，并给出 95%置信区间；采用 Welch t 检验和 Holm-Bonferroni 校正（显著性水平 $p_{\text{sig}} = 0.05$）。评估涵盖领域基准、70 亿参数基础模型和消融研究（完整详情见第 G 节至第 G.3 节）。

5.1. 基因组学（QA-BPE-seq）

数据： 150 bp 双端测序读段（ARTsimulator（Huang 等人，2012），30× 覆盖度，错误率加倍）、GRCh38 参考基因组、GIAB HG002 真值集（Zook 等人，2016年）、CAMI II 宏基因组（Sczyrba 等人，2017）。详情见第 G 节。

基线： 我们将以下方法作为对比：（i）通用分词器——（BPE、SentencePiece（Kudo 和 Richardson，2018）、WordPiece）；（ii）鲁棒性增强方法（BPE-dropout（Provilkov 等人，2020））；（iii）字节级模型（ByT5（Xue 等人，2022）、CANINE（Clark 等人，2021年））；（iv）领域标准 k-mer（6-mer DNABERT（Ji 等人，2021年））；以及（v）神经网络方法（CharFormer（Tay 等人，2022））。

表 *1.* 基因组分词的下游任务性能。数值为 $n = 10$ 次运行的均值及 95%置信区间。时间：相对墙钟时间（BPE=10.0×）。

方法	变体F1	分类群 F1	重构	Time
标准字节对编码	.824±.004	.856±.005	.317±.010	10.0
SentencePiece	.837±.004	.872±.005	.301±.009	10.1
WordPiece	.829±.005	.863±.006	.308±.011	10.0
BPE-dropout	.841±.004	.878±.005	.295±.009	10.2
ByT5	.812±.006	.845±.007	.338±.012	25.3
CANINE	.818±.005	.852±.006	.325±.011	22.7
DNABERT-k	.851±.003	.889±.004	.287±.008	9.8
CharFormer	.856±.003	.893±.004	.279±.008	10.4
QA-BPE-seq	**.891±.004**	**.917±.003**	**.241±.007**	**10.2**

表*2*。 QA-BPE-seq 的消融研究（变体F1）。数值为 $n = 10$ 次运行的均值和 95%置信区间。*

配置	变体F1	Δ(%)
QA-BPE-seq（完整）	**.891±.004**	—
不使用强化学习（贪心 w_{ab}）	.862±.005	−3.3
无质量（$R_Q = 0$）	.825±.004	−7.4
无信息奖励（$R_I = 0$）	.872±.005	−2.1
无自适应参数	.857±.006	−3.8
不使用 R_{bio}	.885±.004	−0.7
QualTok（基线）	.840±.005	−5.7

* “不使用强化学习（贪心 w_{ab}）”使用完整的 QA-Token 合并分数和学习得到的 α，但不进行强化学习策略优化，而是贪心地选择合并项。“QualTok（基线）”还固定了自适应参数（α=0.5，均匀权重）。

质量设计： 带位置衰减的 Phred 分数、几何均值聚合、学习得到的 $\alpha = 0.72 \pm 0.03$,$\beta_{\text{pos}} = 0.014 \pm 0.002$。

评估：（i）通过 Transformer 模型进行变异检测，该模型以词元嵌入作为特征并预测变异，使用 `hap.py` 与 GIAB 真值集进行评估；（ii）分类学分类（6 层 Transformer）；（iii）序列重建（自编码器），遵循既定的基准测试协议（Rumpf 等人，2023年）。表1 显示 QA-BPE-seq 的表现优于全部基线（$p < 0.001$）。

关键洞见：（i）QA-BPE-seq 在变异检测中实现了 6.7 个百分点的 F1 提升（0.891，而 BPE 为 0.824）。（ii）字节级模型出现灾难性失败（慢 2.5×倍，准确率低 7–9 %）。（iii）涌现词汇表在高质量区域无需显式监督即可与生物学单元（密码子、基序）对齐（词汇表分析见第 G 节）。

表 3. QAT-QF 消融研究（收益预测准确率 % 和夏普比率）。$n = 10$ 次运行的均值及 95%置信区间。*

变体	收益预测（%）	Sharpe
完整模型	**68.3±0.5**	**1.72±0.07**
不含质量（$R_Q = 0$）	64.2±0.6	1.56±0.08
不含信息（$R_I = 0$）	65.1±0.5	1.61±0.07
无预测功率 ($R_P = 0$)	63.9±0.6	1.49±0.09
无复杂度 ($R_C = 0$)	66.8±0.4	1.73±0.06
固定 α	65.4±0.5	1.65±0.07
固定 γ	64.9±0.5	1.59±0.08
QualTok-QF（基线）	64.8±0.6	1.58±0.08

* “QualTok-QF（基线）”采用简化的质量感知合并评分，使用固定的 α=0.5 和均匀权重，不进行强化学习策略优化或自适应参数学习。

5.2. 量化金融（QAT-QF）

数据集：我们使用来自 LOBSTER（Huang& Polak, 2011）的 BTC/USD 交易对高频限价订单簿（LOB）数据，具体为 2023年第一季度的 10 档重建快照。数据按时间顺序划分为 70% 训练集、15% 验证集和 15% 测试集。原子元素定义为连续 5 个限价订单簿事件的序列，编码为元组（Δ中间价 Δ价差，成交量_失衡，事件_类型 Δt），离散化方式为：价格变动分为 10 个分箱（±5 个跳动点），价差分为 10 个分箱，成交量失衡分为 5 个有符号分箱，事件类型为分类变量（成交/撤单/限价订单），时间间隔分为 5 个对数间隔分箱，从而得到 $|\Sigma| = 7{,}500$ 个原子符号（见第 D 节）。

基线： QAT-QF 与多种适用于金融时间序列的分词和离散化方法进行了基准比较。

- **通用方法：** 标准字节对编码、SentencePiece（Unigram LM 模式）和 BPE-dropout（Provilkov 等人, 2020）用于评估鲁棒性。
- **时间序列专用方法：** 符号聚合近似（SAX）（Lin 等人, 2003）（PAA=16，字母表大小=8）以及 SFA 符号袋（BOSS）（Schäfer, 2015年），两者都广泛用于符号化时间序列表示。

子词模型的目标词汇表大小为 16,000。

评估： 我们评估：(i) 收益预测准确率（5 分钟中间价收益率方向），(ii) 波动率预测均方根误差（5 分钟已实现波动率），(iii) 市场状态识别（2 状态 GARCH-HMM 分类），以及 (iv) 交易表现（夏普比率（Sharpe, 1994），并计入 5 个基点的交易成本）。模型使用 2 层长短期记忆网络（128 个隐藏单元）和近端策略优化智能体（Deng 等人, 2016年）。实现细节请参见第 D 节和第 H.4 节。

结果：表4 给出了在 $n = 10$ 运行次数上取平均后的结果。QAT-QF在全部金融任务上均提升了性能（$p < 0.01$，经Holm-Bonferroni校正）。交易智能体取得的夏普比率为 1.72 ± 0.07，相比1.32 ± 0.05（标准字节对编码，提升30%）。消融分析见表3。

表4. 金融领域词元化的下游任务性能。数值为 $n = 10$ 次运行的均值及95%置信区间。时间：每轮分钟数。

方法	收益率（%）	Vol.	市场状态	Sharpe	Time
BPE	61.2±0.5	.014±.001	73.5±0.6	1.32±.05	15.0
SAX	58.9±0.6	.014±.001	75.2±0.5	1.29±.06	14.5
BOSS	62.3±0.4	.013±.001	78.4±0.4	1.45±.05	14.8
QAT-QF	**68.3±0.5**	**.010±.001**	**86.4±0.3**	**1.72±.07**	**15.2**

6. 基础模型验证

我们在领域基准上验证 QA-Token（第 5 节），现在进一步在基础模型规模上进行评估。我们重新训练基因组学和金融领域的最先进基础模型，以证明质量感知分词能够改善大型模型从噪声语料库中的学习方式，从而不同于传统的基于频率的方法。

6.1. 宏基因组学基础模型：METAGENE-1 7B

设置： 重新进行分词的 METAGENE-1（Liu 等人, 2025）（70 亿参数，1.7 万亿碱基对），架构和超参数保持不变，对比 BPE 与 QA-BPE-seq。

质量感知设计： 分词器使用基因组质量评估指标（公式 13，第 D 节）在 20 亿碱基对（语料库的 0.12%）上进行训练，结合了 (i) 基于 Phred 的质量分数，(ii) 通过 k-mer 分析得出的保守性分数，(iii) GC 含量偏差指标，以及 (iv) 二级结构预测置信度。学得的 $\beta_{\text{pos}} = 0.014$ 捕获了位置特异性质量衰减（实现方法见第 H.1 节）。

训练预算： 两个模型处理相同的原始数据量（1.7 万亿碱基对）。词元减少 15% 意味着 QA-BPE-seq 能以更少的优化步数完成训练轮数，同时保持相同的原始数据接触量。步数匹配实验（优化步数相同，此时 QA-BPE-seq 每步处理的原始数据多 17.6%）显示出一致的改进（见第 G 节）。

病原体检测： QA-Token 取得了最先进的 94.53 马修斯相关系数，比原始 METAGENE-1 高出 1.57 分（$p < 0.001$）。五个子任务均有稳定提升，展现出其稳健性。在质量感知最为关键的高度降解宏基因组样本上，任务 2 的提升幅度最大（+2.04 马修斯相关系数），验证了

表 5. 病原体检测基准（马修斯相关系数）。QA-Token 达到最先进水平。

模型	T-1	T-2	T-3	T-4	T-5	平均
DNABERT	82.2	81.4	83.3	84.6	82.9	82.9
DNABERT-2	86.7	86.9	88.3	89.8	87.9	87.9
DNABERT-S	85.4	85.2	89.0	88.4	86.0	87.0
NT-2.5B-M	83.8	83.5	82.5	79.9	81.4	82.4
NT-2.5B-1k	77.5	80.4	79.8	78.4	79.0	79.0
HyenaDNA	78.7	79.1	80.4	81.2	79.9	79.9
METAGENE-1	92.1	90.9	93.7	95.1	94.0	93.0
+QA-Token	**93.8**	**93.0**	**95.1**	**96.2**	**94.5**	**94.5**
Δ	+1.7	+2.0	+1.4	+1.1	+0.6	+1.6

表6. 基因组理解评估（GUE）：多物种基准测试。

Task	META-1	QA-Token	Δ	p
调控元件				
TF-Mouse（马修斯相关系数）	71.4±0.8	**72.8±0.7**	+1.4	.002
TF-HUMAN（马修斯相关系数）	68.3±0.9	**69.9±0.8**	+1.6	.001
启动子（马修斯相关系数）	82.3±0.5	**85.5±0.4**	+3.2	<.001
增强子（曲线下面积）	.876±.01	**2.892±.010**	+.016	.003
表观遗传学				
H3K4三甲基化（马修斯相关系数）	65.2±0.6	**66.8±0.5**	+1.6	.002
H3K27ac（马修斯相关系数）	66.8±0.7	**68.2±0.6**	+1.4	.003
甲基化（曲线下面积）	.823±.015	**.841±.013**	+.018	.004
结构				
剪接位点（F1 分数）	87.8±0.4	**89.5±0.3**	+1.7	<.001
RNA结构	72.1±0.8	**73.9±0.7**	+1.8	.002
变异				
COVID（F1 分数）	72.5±0.6	**73.3±0.5**	+0.8	.018
SNP效应	.684±.021。	**712±。018**	+.028	.001
胜率	46.4%	**57.1%**	**+10.7%**	—
效率	370B	**315B**	**−15%**	—

我们的理论框架。

GUE 结果： QA-Token 在所有类别中均提升了性能（最大提升： +3.2 马修斯相关系数，见启动子检测）。词元减少 15% 且性能有所提升，表明质量感知合并具有语义连贯性。

6.2. 金融时间序列基础模型

设置： 受 TimesFM（Das 等人，2024年）和 Chronos（Ansari 等人，2024年）启发的 12 亿参数模型（24 层，维度为 2048），使用 QAT-QF 处理噪声。

训练语料库： 我们在覆盖 5000 亿个时间序列观测值的数据上进行训练，这些数据包括：(i) 高频订单簿数据（40%，涵盖 50 种高流动性资产、时间跨度为 5 年、毫秒级分辨率），(ii) 日频 OHLCV 数据（30%，涵盖主要指数、时间跨度为 20 年），(iii) 宏观经济指标（20%，涵盖二十国集团 30 年的数据），以及 (iv) 另类数据（10%，包括情绪评分、期权流量和 ETF 组成）。

表 7. 金融基础模型评估（100 个测试回合）。

Task	零样本 BPE	零样本 量化感知训练	零样本 Δ	少样本 BPE	少样本 量化感知训练	少样本 Δ
价格预测						
方向 5 分钟	52.3	**58.7**	+12	61.2	**68.3**	+12
方向 1小时	51.8	**57.2**	+10	59.4	**65.8**	+11
方向 1天	50.9	**54.6**	+7	56.7	**61.2**	+8
收益均方误差	1.00	**0.81**	−19	0.72	**0.60**	−18
波动率						
波动率均方根误差	.018	**.014**	−23	.013	**.010**	−27
GARCH 估计值	.156。	**118**	−24	.098	**.071**	−28
波动率状态	71.2	**79.8**	+12	82.3	**88.4**	+7
微观结构						
价差	.023	**.019**	−20	.018	**.013**	−25
成交量	31.2	**24.8**	−21	22.6	**17.3**	−24
订单流	.412	**.523**	+27	.567	**.681**	+20
Risk						
市场状态 F1	.673	**.751**	+12	.798	**.856**	+7
回撤	.682	**.743**	+9	.761	**.812**	+7
尾部风险	.412	**.486**	+18	.523	**.598**	+14
跨资产						
相关性预测	.623	**.694**	+11	.712	**.768**	+8
领先-滞后	58.3	**64.7**	+11	67.2	**73.1**	+9
轮动	1.23	**1.41**	+15	1.52	**1.72**	+13
平均值 Δ	—	—	**+16%**	—	—	**+13%**

质量感知设计： QAT-QF 采用综合市场质量指标（公式 14，第 D 节），结合流动性、信号、稳定性和信息质量等维度。学习得到的权重 w_k 会适应不同的市场状态，并使用 $\beta_{\text{vol}} = 0.50 \pm 0.05$ 进行波动率缩放（完整参数设置请参见第 H.2 节）。

评估指标：方向 = 方向准确率（%）；收益均方误差 = 收益预测均方误差（归一化至 BPE=1.0）；波动率均方根误差 = 波动率预测均方根误差；订单流 = 订单失衡预测 R^2；市场状态 F1 =市场状态分类 F1 分数；尾部风险 = VaR 超限预测 F1 分数；轮动 =行业轮动策略夏普比率。

财务结果： QAT-QF 实现了 7.3–27.0% 的零样本改进，在波动率/微观结构任务中提升最大。订单流失衡（+27.0%），市场状态检测（+11.6% F1 分数）证明了 QA-Token 的噪声过滤能力，与我们的信息论最优性结果（定理C.13）一致。实现细节见第 F–第 G.3节。

计算成本： QA-Token 的词汇表构建需要 50–60 个图形处理器小时，而标准字节对编码只需几分钟。不过，这项一次性成本可以分摊到数十亿次推理操作中：词汇表构建完成后不会带来额外的推理开销——分词速度与 BPE 相同（∼10毫秒/序列），因为质量指标仅在构建过程中使用。这种效率适用于高性能计算系统和存储内处理架构（Mansouri Ghiasi 等，2022；Ghiasi 等，2022；2023年；Mansouri Ghiasi 等，2023年；Ghiasi 等，2024年）。对于基础模型

其中分词仅执行一次，却会影响数十亿次推理操作，因此额外的前期成本可通过显著的长期收益得到合理补偿；对于小规模应用或干净数据集，标准字节对编码可能仍然更加实用。

7. 结论

QA-Token 将分词从频率统计扩展到质量驱动的词汇表构建，解决了处理含噪真实世界数据时的局限性。我们提出了：(i) 双层优化及其 NP难性证明（定理3.2，第 C.5 节）；(ii) 马尔可夫决策过程形式化，达到 $(1-1/e)$-近似（定理E.4、定理 E.5、第 E.7 节）；(iii) Gumbel-Softmax 支持端到端学习（定理 C.8，第 C.5 节）。评估结果显示出持续的改进：(1) 基因组学——F1 分数提高 6.7 个百分点，病原体检测的马修斯相关系数达到 94.53；(2) 金融——夏普比率提高 30%；(3) 基础模型实现了新的基准测试结果（分析见第 G 节至第 G.3 节）。随着生物序列档案扩展到拍碱基规模（Karasikov等人，2025），以及变异预测方法实现前所未有的准确率（Avsec等人，2026），质量感知分词对于弥合数据可用性与基础模型可用性之间的差距变得至关重要。

7.1. 科学与经济影响

QA-Token 使人们能够利用此前被认为无法使用的海量噪声数据集，从根本上拓展了基础模型训练的数据边界。

基因组学的科学加速。 序列读段档案（SRA）包含超过 67 拍碱基的公开基因组数据——相当于阅读人类基因组 2200 万次——然而，由于质量异质性（Leinonen等人，2011），其中相当大一部分仍未得到充分利用。近期基础设施的进步使这些拍字节级档案库能够以经济的成本进行全文检索（Karasikov等人，2025年），而 AlphaGenome 等先进方法如今可实现对调控变异效应的精准预测（Avsec等人，2026）。然而，数据可访问性与数据可用性之间的鸿沟在基础模型训练中依然存在：标准分词方法要么完全丢弃低质量读段，要么将测序错误传播到学习得到的表征中。QA-Token 通过实现质量感知分词来弥合这一鸿沟，从而利用全部可用序列数据。我们展示了三个关键应用：（1）大流行监测——用于病原体监测的环境样本中，40–60% 的内容因污染和测序错误而含有噪声；QA-Token 可直接利用此类含噪宏基因组数据进行训练（Gollwitzer等人，2023b；a；2025a），在病原体检测中取得了 94.53 的马修斯相关系数，并借助此前无法使用的环境样本，实现全球大流行实时监测。（2）药物发现——长读长测序-用于检测结构变异时的错误率为 10–15%；我们在变异检测中将 F1 分数提高了 6.7 个百分点，从而加快了从复杂基因组重排中识别药物靶点的进程，并与调控变异预测方面的进展相辅相成（Avsec等人，2026）。（3）进化生物学——古DNA严重降解，损伤率达 >50%；质量感知分词在过滤损伤的同时保留真实的古序列，让研究人员得以从此前无法分析的样本中获得进化学见解。

金融领域的经济影响。 全球金融市场每天产生 5TB 数据，其中 40% 含有由市场碎片化和延迟造成的微观结构噪声；目前的方法需要昂贵的数据清洗基础设施，每年成本高达数百万。QA-Token 能带来可量化的经济价值：(1)算法交易——夏普比率提高 30%，意味着大型基金可获得数十亿美元的额外回报；订单流预测准确度提高 27%，每天可节省价值数百万的基点级执行成本。(2) 风险管理——尾部风险估计能力提高 18%，本可避免市场崩盘期间数十亿美元的损失；市场状态检测能力提高 11.6%，使投资组合能够更快地重新平衡。(3)普及化——规模较小的机构如今无需昂贵的数据清洗基础设施也能参与竞争，从而降低了采用量化交易策略的准入门槛。

更广泛的社会影响。 除基因组学和金融领域外，QA-Token还具有以下潜在应用： 医疗保健——医院每天都会产生数太字节的含噪医学数据；QA-Token 支持利用带有伪影的真实临床数据进行训练，有望提高诊断准确率和治疗建议的质量，包括在癌症治疗优化方面的应用（Gollwitzer等人，2025b）。气候科学——卫星影像常受云覆盖和大气干扰影响；QA-Token允许直接利用部分受损的地球观测数据进行训练，从而加快气候监测和预测能力的发展。 基础设施监测——传感器网络会产生数拍字节的数据，且故障频发；质量感知词元化即使在传感器退化的情况下，也能实现稳健的异常检测，可用于智慧城市应用和工业物联网。

7.2. 局限性与未来工作

局限性： (1) QA-Token需要领域特定的质量信号；对于尚无成熟评估指标的领域，需要进行定制设计。(2) 词汇表构建的开销限制了开发过程中的快速迭代。(3) 有效的质量函数设计需要领域知识，不过自适应学习降低了对初始选择的敏感性。

未来方向： （1）基于数据统计（局部熵、一致性）的通用质量度量。（2）针对流式数据的在线适配。（3）扩展至视觉——

语言和音频文本等多模态领域。（4）通过蒸馏和剪枝提升效率。

影响声明

公共序列存储库目前存有超过 67 拍字节对的原始测序数据，欧洲核苷酸档案库的数据量约每 45 个月翻一番（Karasikov等人，2025）。近期的进展使这些拍字节级档案库能够以每次查询每百万碱基对低至 0.74 美元的成本进行全文检索，表明大规模序列分析的基础设施正迅速成熟。然而，由于质量异质性，其中相当大一部分数据仍未充分用于基础模型训练。QA-Token 弥合了数据可访问性与数据可用性之间的差距，实现质量感知分词，从而利用全部可用序列数据进行基础模型训练。

基因组学。对于含有 40–60% 噪声的环境样本，我们在病原体检测任务中取得了 94.53 的马修斯相关系数，使得利用以往无法使用的宏基因组数据进行实时大流行病监测成为可能。我们的变异检测 F1 分数提升了 6.7 个百分点，可加快从复杂基因组重排中识别药物靶点的速度，即使测序错误率高达 10–15%。这项技术理论上也可能被滥用于生物监测；我们设计 QA-Token 的初衷是用于研究，并配有标准的机构防护措施。

金融。 全球金融市场每天产生 5TB 数据，其中 40% 含有微观结构噪声。夏普比率提升 30% 可转化为算法交易中可量化的收益，而订单流预测准确度提高 27% 可降低执行成本。交易表现的提升引发了对市场公平性的担忧；QA-Token 在现有监管框架内提供渐进式改进。

资源。 构建词汇表需要 50–60 GPU 小时，成本远低于基础模型训练，因此计算预算有限的研究人员也能使用 QA-Token。高度压缩的质量感知词汇表便于移作进一步分析之用。

可复现性声明

本文及附录提供了全面的详细信息。

理论贡献：所有定理和命题均包含完整证明（第C.5节、第C.2节、第C.5节、第C.5节、第C.9节），并附有明确假设（第C.6节）和收敛保证（第E.4节、第E节）。

算法：强化学习策略优化的完整伪代码（算法1）、自适应参数学习（算法2）和最终词汇表构建（算法3）见第E节。

实现： 领域特定质量指标及其精确公式（第2节、第D节）、所有模型的超参数（第 H.1节、第 H.2节）以及计算需求（第 G.3节）均有完整说明。

实验方案： 第 5 节和第 G节详细介绍了统计方法，包括 10 次独立试验、95% 置信区间、采用 Holm-Bonferroni 校正的 Welch t 检验，以及效应量。数据集规范、预处理步骤和评估指标见第 G节至第 I.2节。

基线：九种基线方法的实现细节和超参数详见第5 节和第I.4节。

代码发布：我们将提供一个 GitHub 仓库，其中包含所有源代码、训练好的模型和用于复现实验结果的脚本。

利益冲突声明

A.E.G. 和 D.d.G. 是 Anto Biosciences（YC F25）的联合创始人和股东。

D.A.S.、P.L. 和 A.N.d.l.C. 声明不存在利益冲突。

参考文献

Alemi, A. A., Fischer, I., Dillon, J. V. 和 Murphy, K. 深度变分信息瓶颈。载于国际学习表征会议（*ICLR*），2017年5月。

Andersen, T. G., Bollerslev, T., Diebold, F. X. 和 Labys, P. 已实现汇率波动率的分布。美国统计学会会刊，96(453)：42–55，2001年。

Ansari, A. F., Stella, L., Turkmen, C., 张, X., Mer- cado, P., 沈, H., Shchur, O., Rangapuram, S. S., Pineda Arango, S., Kapoor, S. 等。Chronos：学习时间序列的语言。*arXiv* 预印本 *arXiv:2403.07815*，2024年。

Avsec, Ž., Latysheva, N., Cheng, J. 等。借助 AlphaGeno me 推进调控变异效应预测。《自然》，649：1206–1218，2026年1月。doi：10.1038/ s41586-025-10014-0。

Baldwin, T., Cook, P., Lui, M., MacKinlay, A., 和 Wang, L. 噪声文本分析。载于澳大拉西亚语言技术协会研讨会论文集 *2013*年，第 1–10 页，2013年。

Barbieri, F., Camacho-Collados, J., Ronzano, F., Espinosa-Anke, L., Ballesteros, M., Basile, V., Patti, V., 和 Sag- gion, H. SemEval 2018 任务 2：多语言表情符号预测。载于第十二届国际语义评测研讨会论文集，第 24–33 页，2018。

Barbieri, F., Camacho-Collados, J., Espinosa-Anke, L., 和 Neves, L. TweetEval：面向推文分类的统一基准测试与比较评估。载于*EMNLP* 论文集论文集，2020。

Basile, V., Bosco, C., Fersini, E., Nozza, D., Patti, V., Rangel Pardo, F. M., Rosso, P., 和 Sanguinetti, M. SemEval-2019 任务 5：Twitter 上针对移民和女性的多语言仇恨言论检测。载于第 *13* 届语义评估国际研讨会论文集，第 54–63 页，美国明尼苏达州明尼阿波利斯，2019。计算语言学协会。doi：10.18653/v1/S19-2007。

Bello, I., Pham, H., Le, Q. V., Norouzi, M., 和 Bengio, S. 结合强化学习的神经组合优化。载于国际学习表征会议，2016年。

Bhandari, J. 和 Russo, D. 策略梯度方法的全局最优性保证。运筹学，69(6): 1744–1767，2021年12月。doi: 10.1287/opre.2021.0014。

Bojanowski, P.、Grave, E.、Joulin, A. 和 Mikolov, T. 利用子词信息丰富词向量。载于计算语言学协会会刊，第5卷，第135–146页，2017。

Bolte, J.、Le, Q.-T.、Pauwels, E. 和 Vaiter, S. 双层规划的几何与计算难度。数学规划，2024年。doi: 10.1007/ s10107-025-02229-w。

Borkar, V. S. 随机逼近：动力系统视角。印度斯坦图书社，2009年。

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. 等。《语言模型是少样本学习者》。载于神经信息处理系统进展，第 33 卷，第 1877–1901 页，2020。

Cen, S. 和 Chi, Y.《强化学习、博弈与控制中策略梯度方法的全局收敛》。*arXiv* 预印本 *arXiv:2310.05230*，2023年。

Chai, B. Y., Wang, Z., 和 Sachan, M.《分词的诅咒》。*arXiv* 预印本 *arXiv:2402.07831*，2024年。

Church, K. W. 和 Hanks, P. 词语联想规范、互信息与词典编纂学。计算语言学，16(1):22–29，1990年。

Clark, J. H.、Garcia, D.、Botha, J.、Lee, K.、Luong, M.- T. 和 Le, Q. V. CANINE：为语言表示预训练高效的无分词编码器。载于第*59*届计算语言学协会年会暨第*11*届自然语言处理国际联合会议论文集（第*1*卷：长篇论文），第2647–2661页，2021年。

Das, A.、Kong, W.、Leach, A.、Sen, R. 和 Yu, R. TimesFM：用于时间序列预测的仅解码器基础模型。*arXiv* 预印本 *arXiv:2310.10688*，2024年。

Dempe, S. 《双层优化：理论、算法与应用》，第 161 卷，丛书《*Springer* 优化及其应用》。Springer，德国柏林，2020。doi: 10.1007/978-3-030-33566-3。

Deng, Y., Bao, F., Kong, Y., Ren, Z., andDai, Q. 用于金融信号表示和交易的深度直接强化学习。*IEEE* 神经网络与学习系统汇刊，28(3):653–664，2016年。

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert：用于语言理解的深度双向变换器预训练。载于计算语言学协会北美分会 *2019* 年会议论文集：人类语言技术，第 *1* 卷（长篇与短篇论文），第 4171–4186 页，2019。

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 使用 Phred 对自动测序仪轨迹进行碱基判读。I. 准确性评估。基因组研究，8(3):175–185，1998。

Finn, C., Abbeel, P. 和 Levine, S. 面向深度网络快速适应的模型无关元学习。载于国际机器学习会议，第 1126–1135 页。PMLR，2017。

Gençay, R., Selçuk, F. 和 Whitcher, B. 《金融与经济学中的小波及其他滤波方法导论》。爱思唯尔，圣迭戈，2001年。

Ghiasi, N. M., Park, J., Mustafa, H., Kim, J., Olgun, A., Gollwitzer, A., Cali, D. S., Firtina, C., Mao, H., Alserr, N. A. 等。Genstore：面向高性能和节能基因组分析的基因组数据存内过滤。载于*2022*年*IEEE*计算机学会*VLSI*年度研讨会（*ISVLSI*），第 283–287 页。IEEE，2022年。

Ghiasi, N. M., Sadrosadati, M., Mustafa, H., Gollwitzer, A., Firtina, C., Eudine, J., Ma, H., Lindegger, J., Cavlak, M. B., Alser, M., 等。Metastore：通过存内计算实现高性能宏基因组分析。*arXiv* 预印本 *arXiv:2311.12527*，2023年。

Ghiasi, N. M., Sadrosadati, M., Mustafa, H., Gollwitzer, A., Firtina, C., Eudine, J., Mao, H., Lindegger, J., Cavlak, M. B., Alser, M., 等。Megis：通过存内处理实现高性能、节能且低成本的宏基因组分析。发表于*2024*年 *ACM/IEEE* 第 *51* 届计算机体系结构年度国际研讨会（*ISCA*），第 660–677 页。IEEE，2024年。

Gollwitzer, A., Alser, M., Bergtholdt, J., Lindegger, J., Rumpf, M.-D., Firtina, C., Mangul, S. 和 Mutlu, O. Metafast：通过种子计数和编辑距离近似实现快速宏基因组分类。*arXiv*，第 2311–02029 页，2023a。

Gollwitzer, A. E., Alser, M., Bergtholdt, J., Lindegger, J., Rumpf, M.-D., Firtina, C., Mangul, S. 和 Mutlu, O. Metatrinity：通过种子计数和编辑距离近似实现快速宏基因组分类。*arXiv*预印本 *arXiv:2311.02029*，2023b。

Gollwitzer, A. E., Subramanian, D. A., Tucker, I. 和 Traverso, G. Metaomics-10t：解锁微生物生态系统因果建模的基础数据集。载于*NeurIPS 2025*人工智能促进科学研讨会，2025a。

Gollwitzer, A. E., Subramanian, D. A., Tucker, I. 和 Traverso, G. 掌控进化博弈：癌症治疗中治疗耐药性的层级控制。发表于*NeurIPS 2025*人工智能促进科学研讨会，2025b。

Golovin, D. 和 Krause, A. 自适应次模性：主动学习和随机优化中的理论与应用。发表于第*24*届神经信息处理系统国际会议论文集，第2675–2683页，2011年。

Grne, C. 和 Wulf, L. 双层优化和鲁棒优化中许多自然问题的多项式层级完备性。发表于整数规划与组合优化会议，2023年。doi: 10.1007/978-3-031-93112-319._

Hamilton, J. D. 非平稳时间序列与商业周期经济分析的新方法。《计量经济学》：计量经济学会期刊，第357–384页，1989年。

Han, B., Cook, P., 和 Baldwin, T. 短文本消息的词汇规范化：Makn sens a #twitter。载于计算语言学协会第*51*届年会论文集（第*1*卷：长篇论文），第368–378页，2013年。

Hansen, P. R. 和 Lunde, A. 已实现方差与市场微观结构噪声。《商业与经济统计学杂志》，24(2):127–161，2006。

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Aken, B., Barrell, D., Mudge, J. M., FRecognision, E., GCoil, A., LNCipedia, A. 等。Gencode：ENCODE 项目的人类参考基因组注释。基因组研究，22(9):1760–1774，2012。Hasbrouck, J. 衡量股票交易的信息含量。《金融杂志》，46(1):179–207，1991。Heafield, K. Kenlm：更快、更精简的语言模型查询。载于第六届统计机器翻译研讨会论文集，第 187–197 页。计算语言学协会，2011。

Heinzinger, M., Elnaggar, A., Wang, Y., Dallago, C., Neettiyath, U., 和 Rost, B. 通过迁移学习蛋白质序列，建模生命语言的各个方面。*BMC*生物信息学，20(1):1–17，2019。

Hinton, G., Vinyals, O., 和 Dean, J. 提炼神经网络中的知识，2015年。

Huang, R. 和 Polak, T. Lobster：限价订单簿重建系统。可从*SSRN 1920143*获取，2011。

Huang, W.、Li, L.、Myers, J. R. 和 Marth, G.T. Art：新一代测序读段模拟器。*Bioinformatics*，28（4）：593–594，2012年。

Jang, E.、Gu, S. 和 Poole, B. 使用 Gumbel-Softmax 进行分类重参数化。在《国际学习表征会议》上，2017。

Ji, Y.、Zhou, Z.、Liu, H. 和 Davuluri, R. V. DNABERT：用于 DNA 基因组语言的基于 Transformer 模型的预训练双向编码器表征。*Bioinformatics*，37（15）：2112–2120，2021年。

Jones, D. R.、Schonlau, M. 和 Welch, W. J. 昂贵黑盒函数的高效全局优化。*Journalof Global optimization*，13（4）：455–492，1998年。

J.P. Morgan。RiskMetrics 技术文档。技术报告，J.P.Morgan/Reuters，1996年。

Karasikov, M., Mustafa, H., Danciu, D., Bosshard, L., Zimmermann, M., Schütze, K., Kahles, A. 和 Rätsch, G. 面向 PB 级序列库的高效、准确搜索。《自然》647:1036–1044，2025年。doi: 10.1038/s41586-025-09603-w。

Karp, R. M. 组合问题之间的可归约性。载于《计算机计算复杂性》，第 85–103 页。Springer，1972年。

Kingma, D. P. 和 Ba, J. Adam：一种随机优化方法。*arXiv* 预印本 *arXiv:1412.6980*，2014。

Kudo, T. 子词正则化：利用多个子词候选改进神经网络翻译模型。载于第 *56* 届计算语言学协会年会论文集（第 *1* 卷：长篇论文），第 66–75 页，2018。

Kudo, T. 和 Richardson, J. SentencePiece：一种简单且与语言无关的子词分词器和反分词器，用于神经文本处理。载于*2018*年自然语言处理经验方法会议论文集：系统演示，第 66–71 页，2018 年。

Leinonen, R.、Sugawara, H.、Shumway, M. 和国际核苷酸序列数据库协作组。序列读档案。《核酸研究》，39（增刊_1）：D19–D21，2011 年。

Li, J.、Park, Y.-B.、Song, Y.-S. 和 Park, S.-K. 针对各种韩语自然语言处理任务的分词策略实证研究。载于第 *12* 届语言资源与评估会议论文集，第 6813–6819 页，2020 年。

Libovick'y, J. 和 Sachan, M. 用于提升大语言模型性能的语义切分。载于计算语言学协会论文集：*ACL 2024* 年发现，第 4930–4945 页，2024 年。

Lin, H. 和 Bilmes, J. 一类用于文档摘要的次模函数。载于计算语言学协会第 *49* 届年会论文集：人类语言技术，第 510–520 页。计算语言学协会，2011。

Lin, J.、Keogh, E.、Lonardi, S. 和 Chiu, B. 时间序列的符号表示及其对流算法的影响。载于第 *8* 届 *ACM SIGMOD* 数据挖掘与知识发现研究问题研讨会论文集，第 2–11 页，2003。

Liu, O. 等。METAGENE-1：用于大流行监测的宏基因组基础模型。*arXiv* 预印本 *arXiv:2501.02045*，2025 年 1 月。

Maddison, C. J.、Mnih, A. 和 Teh, Y. W. Concrete 分布：离散随机变量的连续松弛。载于国际学习表征会议，2017。

Madhavan, A. 市场微观结构：综述。金融市场杂志，3(3):205–258，2000。

Mansouri Ghiasi, N., Park, J., Mustafa, H., Kim, J., Olgun, A., Gollwitzer, A., Senol Cali, D., Firtina, C., Mao, H., Al-madhoun Alserr, N., 等。Genstore：一种用于基因组序列分析的高性能存储内处理系统。载于第 *27* 届 *ACM* 编程语言与操作系统架构支持国际会议论文集，第 635–654 页，2022年。

Mansouri Ghiasi, N., Sadrosadati, M., Mustafa, H., Goll-witzer, A., Firtina, C., Eudine, J., Ma, H., Lindegger, J., Banu Cavlak, M., Alser, M., 等。Metastore：通过存储内计算实现高性能宏基因组分析。*arXiv* 电子预印本，第 arXiv–2311 页，2023年。

Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., 和 Cherry, C. SemEval-2016 任务 6：检测推文中的立场。载于第 *10* 届语义评估国际研讨会论文集（*SemEval-2016*），第 31–41 页，2016年。

Mohammad, S., Bravo-Marquez, F., Salameh, M., 和 Kir-itchenko, S. SemEval-2018 任务 1：推文中的情感。载于第 *12* 届语义评估国际研讨会论文集，第 1–17 页，2018。

Moody, J. 和 Saffell, M. 交易系统和投资组合的绩效函数与强化学习。预测杂志，20(1):1–18，2001年。

Moody, J. 和 Wu, L. 通过直接强化学习进行交易。载于*IEEE* 神经网络国际会议论文集，第1741–1746页。IEEE，1998。

Nguyen, D. Q., Vu, T. 和 Nguyen, A. T. BERTweet：一种用于英文推文的预训练语言模型。载于*2020*年自然语言处理经验方法会议：系统演示，第9–14页，2020年。

Owen, A.B. 蒙特卡罗理论、方法与示例。斯坦福大学，2013年。可访问 https://artowen.su.domains/mc/。

Provilkov, I., Emelyanenko, D. 和 Voita, E. Bpe-dropout：简单有效的子词正则化。载于计算语言学协会第 *58* 届年会论文集，第 1882–1892 页，2020年。

Ranzato, M., Chopra, S., Auli, M. 和 Zaremba, W. 使用循环神经网络进行序列级训练。载于国际学习表征会议，2015年。

Rissanen, J. 通过最短数据描述进行建模。*Auto matica*，14(5)：465–471，1978年。

Rosenthal, S., Farra, N. 和 Nakov, P. SemEval-2017 任务 4：Twitter 情感分析。载于第 *11* 届语义评估国际研讨会论文集（*SemEval-2017*），第 502–518 页，2017。

Rumpf, M.-D., Alser, M., Gollwitzer, A.E., Lindegger, J., Almadhoun, N., Firtina, C., Mangul, S., and Mutlu, O. Sequencelab：用于比较基因组序列的计算方法综合基准测试。*arXiv* 预印本 *arXiv:2310.16908*，2023年。

Rusu, A. A., Colmenarejo, S. G., Gulcehre, C., Desjardins, G., Kirkpatrick, J., Pascanu, R., Mnih, V., Kavukcuoglu, K., and Hadsell, R. 策略蒸馏，2016年。

Schäfer, P. 噪声存在时，BOSS 关注时间序列分类。《数据挖掘与知识发现》，29(6):1505–1530，2015年。

Schulman, J.、Wolski, F.、Dhariwal, P.、Radford, A. 和 Klimov, O. 近端策略优化算法。载于 *arXiv* 预印本 *arXiv:1707.06347*，2017。

Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dr"oge, J., Gregor, I., Majda, S., Fiedler, J., Dahms, E. 等。《宏基因组解读的关键评估——宏基因组学软件基准测试》。《自然·方法》，14(11):1063–1071，2017。Sennrich, R., Haddow, B. 和 Birch, A. 使用子词单元对罕见词进行神经机器翻译。载于《计算语言学协会第54届年会论文集（第1卷：长篇论文）》，第1715–1725页，2016。Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., 张, M., Li, Y., Wu, Y. 和 Guo, D. Deepseekmath：推动开放语言模型数学推理的极限。《arXiv》预印本 arXiv:2402.03300，2024。Sharpe, W. F. 夏普比率。《投资组合管理杂志》，21(1):49–58，1994。Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. 和 Sirotkin, K. dbSNP：美国国家生物技术信息中心遗传变异数据库。《核酸研究》，29 (1):308–311，2001。Subramanyam, A., Chen, Y. 和 Grossman, R. L. 重新审视规模定律：建模数据质量在语言模型预训练中的作用。《arXiv.org》，2025。doi: 10.48550/arXiv.2510.03313。Sutton, R. S. 和 Barto, A. G.《强化学习：导论》。麻省理工学院出版社，2018。Tay, Y., Tran, V. Q., Ruder, S., Gupta, J., Liu, L., Chung, J., Turner, S., Wang, Z., Williams, D., Casas, D. G. 等。CharFormer：通过基于梯度的子词分词实现快速字符变换器。《arXiv》预印本 arXiv:2106.12672，2022。Tishby, N., Pereira, F. C. 和 Bialek, W. 信息瓶颈方法。载于《第37届阿勒顿通信、控制与计算会议论文集》，第368–377页，1999。Van Hee, C., Lefever, E. 和 Hoste, V. SemEval-2018 任务 3：英语推文反讽检测。载于《论文集》

第*12*届语义评估国际研讨会，第39–50页，2018。

Wenger, A. M.、Peluso, P.、Rowell, W. J.、Chang, P.-C.、Hall, R. J.、Concepcion, G. T.、Ebler, J.、Fungtammasan, A.、Kolesnikov, A.、Olson, N. D. 等。准确的环形共识长读长测序可改善人类基因组的变异检测与组装。《自然·生物技术》，37(10):1155–1162，2019。

Williams, R. J. 用于联结主义强化学习的简单统计梯度跟踪算法。《机器学习》，8(3-4):229–256, 1992.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., 等。《谷歌神经机器翻译系统：弥合人工翻译与机器翻译之间的差距》，2016年。Xue, L., Barua, A., Constant, N., Al-Rfou, R., Narang, S., Kale, M., Roberts, A., 和 Raffel, C. ByT5：迈向无词元的未来——使用预训练的字节到字节模型。计算语言学协会汇刊，10:291–306，2022年。Yue, Y. 等。强化学习是否真的能提升大语言模型超越基础模型的推理能力？*arXiv*预印本 *arXiv:2504.13837*，2025年4月。发表于 NeurIPS 2025（口头报告）、ICML 2025 AI4Math研讨会最佳论文。Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., 和 Kumar, R. SemEval-2019 任务 6：识别和分类社交媒体中的冒犯性语言（OffensEval）。载于第*13*届国际语义评测研讨会论文集，第75–86页，2019年。Zook, J. M., Catoe, D., McDaniel, J., Vang, L., Spies, N., Sidow, A., Weng, Z., 和 Salit, M. 对七个人类基因组进行大规模测序，以表征基准参考材料。《科学数据》，3(1):1–19，2016年。

补充信息

A. 符号说明

为确保清晰严谨，我们在表8中定义数学符号。我们区分原子（不可再分）元素与词元（由原子元素或其他词元构成的序列）。

表 8. 符号说明表

符号	定义
Σ	原子元素的基础字母表（例如字符、DNA碱基）。
s_i	来自 Σ的一个原子元素。
q_i	原子元素 s_i的标量质量分数，其中 $q_i \in [0,1]$。
t, a, b	词元，即由原子元素组成的序列。
V_k	合并步骤 k时的词汇表。
$f(t)$	词元 t 在语料库中的频率。
$\|t\|$	词元 t 所含原子元素的长度。
$n_\sigma(t)$	词元 t中原子元素 $\sigma \in \Sigma$ 的计数。
$H(t)$	词元 t的经验熵： $H(t) = -\sum_\sigma \frac{n_\sigma(t)}{\|t\|} \log \frac{n_\sigma(t)}{\|t\|}$.
$\boldsymbol{q}_t$	词元 t 在多维领域中的质量分数向量。
q_t	词元 t 的聚合标量质量分数，由其组成部分推导得出。
$\bar{q}_{ab}$	组成词元 a的 b平均质量，定义为 $(q_a + q_b)/2$。
α	控制合并分数中对质量敏感度的可学习指数。
w_{ab}	词元对 (a, b) 的质量感知合并评分。
θ_{adapt}	框架中所有可学习的自适应参数组成的向量。
π_{θ_π}	用于选择合并的强化学习策略，由 θ_π 参数化。
L_{task}	下游机器学习任务的损失函数。
$\mathcal{J}(\mathcal{T})$	分词过程的全局目标函数（式 1）。

B. 相关工作

QA-Token 与子词分词、噪声数据处理、用于序列优化的强化学习，以及自适应或可微建模技术等研究相交，并在此基础上进一步拓展。表9提供了比较概述，将 QA-Token 与现有方法进行对比，并突显其独特之处：显式质量整合、基于强化学习的合并优化，以及对分词过程参数的自适应学习。QA-Token 的自适应参数学习，其关键区别在于专注于优化控制分词*过程*本身的参数（例如质量敏感度或奖励组成部分权重），而非仅仅在固定的合并逻辑下调整词汇表内容或分割边界。

表9. QA-Token 与代表性分词方法的比较。

方法	显式质量整合	优化方法	自适应参数（学习到的过程？）	下游感知（通过奖励/损失）	领域噪声模型（显式？）	词汇表 Type
标准字节对编码/WP/SP（Sennrich 等人，2016年；Wu等人，2016年；Kudo 和 Richardson，2018）	No	频率	No	No	No	子词
BPE-dropout（Provilkov 等人，2020）	No	频率+随机	No	No	No	子词
字符/字节模型（Xue 等人，2022；Clark 等人，2021年）	No	不适用（固定）	No	是（通过模型）	隐式	字符/字节
基于梯度（Tay 等人，2022）	No	梯度	是（分段器）	Yes	隐式	字符/子词
语义分词器（Libovick'y & Sachan，2024年）	No	语义+频率	No	间接地	No	子词
QA-Token（我们的方法）	**Yes**	**强化学习（策略）+ 梯度（超参数）**	**是（过程超参数：$\alpha, \lambda_i, w_j, \beta_k$）**	**是（通过强化学习的奖励，$L_{\text{downstream}}$ 超参数）**	**是（通过 Q，R）**	**子词**

注：“自适应参数（学习到的过程？）”指的是学习控制分词*过程*本身的参数（例如 *QA-Token* 的$\alpha, \beta_k, \lambda_i, w_j$），而不仅仅是学习词汇表内容或分割边界。*QA-Token* 使用强化学习来优化合并策略，并使用基于梯度的方法来优化这些过程超参数。

子词分词算法：当前主流范式依赖基于频率的贪心合并过程，典型代表包括 BPE（Sennrich 等人，2016年）、WordPiece（Wu等人，2016年）（它优化数据似然），以及 SentencePiece（Kudo 和 Richardson，2018）（它直接处理原始文本）。尽管这些方法计算高效且普遍有效，但其基本机制忽略了序列质量，这正是我们开展这项工作的主要动机。BPE-dropout

（Provilkov 等人，2020）在合并过程中引入随机性，作为一种正则化手段来增强鲁棒性，但并未使用明确的质量信号。单语语言模型（Kudo，2018）提供了一种概率式替代方案，但仍主要依赖频率和似然目标，缺乏明确的质量感知。

处理嘈杂和领域特定数据：大量研究聚焦于在特定应用领域内对噪声进行建模。在基因组学中，Phred 分数（Ewing 等人，1998）是标准质量指标，专门模型旨在考虑测序错误（Heinzinger 等人，2019）。在自然语言处理中，针对社交媒体文本的大量工作通过文本规范化（Han 等，2013；Li 等人，2020）和显式噪声建模（Baldwin 等人，2013）等技术来应对词汇变异、拼写错误和俚语。金融时间序列分析经常采用滤波方法（Gençay 等人，2001年）、微观结构建模（Madhavan，2000；Hasbrouck，1991）以及机制转换模型（Hamilton，1989）来处理固有噪声与非平稳性。QA-Token 的独特之处在于提供了一个*统一分词框架*，将此类领域特定的质量与噪声考量直接融入词元构建过程本身，而不是仅将噪声作为独立的下游建模挑战来处理。“分词诅咒”这一概念（Chai 等人，2024年）凸显了分词选择对大语言模型稳健性的下游影响，进一步强调了质量感知方法的必要性。

用于序列优化的强化学习：强化学习为不确定条件下的序贯决策提供了稳健的框架（Sutton 和 Barto，2018）。它已成功应用于各种涉及序列的优化问题，包括文本生成（Ranzato 等人，2015年）、组合优化（Bello 等人，2016年）以及金融策略优化（Moody 和 Wu，1998；Moody 和 Saffell，2001年）。我们独特地将分词词汇表构建过程形式化为一个强化学习问题：合并操作构成由学习得到的策略选择的动作，以最大化累积奖励信号；该信号反映词元质量、信息含量、复杂度和估计效用。这种形式化使得我们能够优化与最终分词结果质量相关的复杂且可能不可微的目标函数。奖励本身由自适应学习得到的参数塑造（第4.3节）。

自适应且可微的分词：鉴于静态分词器固有的局限性，研究人员正在探索自适应且可学习的替代方案。基于梯度的方法（Tay 等人，2022）会在训练下游任务的同时端到端地学习分割参数，通常以字符级方式运行。语义分词（Libovick‘y 和 Sachan，2024年）利用词义指导切分过程。QA-Token 以一种独特的方式整合自适应学习：它学习超参数（α，β_k，w_j，λ_i，.....），这些超参数直接控制质量感知的合并决策和强化学习智能体的奖励结构。在优化下游任务损失时，Gumbel-Softmax 松弛（Jang 等人，2017；Maddison 等人，2017）使合并选择能够相对于这些超参数保持可微（通过方程37中定义的复合逻辑值）。这使得基本的*分词逻辑*能够根据观察到的数据属性和任务反馈进行自适应，并与强化学习智能体的策略共同演化。元学习（Finn 等人，2017）为进一步加快在异构数据源（例如，不同的社交媒体平台）之间的适应提供了一种潜在机制；QA-Token 在概念层面探讨了这一机制（见附录E.5）。

本质上，QA-Token 综合了这些相关领域的概念，同时提供了独特的组合：在合并决策中显式整合质量；通过强化学习，利用多维奖励信号优化合并序列；并自适应地学习定义该奖励和合并逻辑的核心过程参数，从而证明其适用于多种噪声领域。

C. 理论框架与证明

C.1. 质量度量证明

命题 C.1(有界　　性与质量函数的连续性)。 所有特定领域的质量函数 $q_t \in [0,1]$ 均　　*e:*

1. 有界：　$0 \le q_t \le 1$对所有词元均成立 *t2.* 连续：关于其自变量利普希茨连续 *3.* 单调：质量随噪声/误差增加而降低

证明。**有界性：**对于基因组学，　$[0,1]$ 中各值的几何平均数仍位于 $[0,1]$内。对于金融，有界分量的凸组合q_k, $_t \in [0,1]$ 与 $\sum_k w_k = 1$ 得到$q_t^{\text{finance}} \in [0,1]$。

利普希茨连续性：对于基因组学（$[\epsilon_Q, 1]^n$上的几何平均数），经由对数变换的链式法则得到 Lipschitz 常数 $L_g = 1/(\sqrt{n} \cdot \epsilon_Q)$。对于金融，利普希茨分量函数的加权和具有 $L_f \leq \max_k L_k$。
单调性：对于任意满足 $\eta(q) \leq q$ 的噪声注入 η，两种聚合方式（几何平均值与算术平均值）均保持排序关系：噪声越大的输入，质量分数越低。 □

C.2. 合并分数推导

引理 C.2（一阶近似）。目标函数 $\mathcal{J}$ 中由合并 $(a, b) \mapsto ab$ 带来的边际增益可作如下分解：

$$\boxed{\Delta\mathcal{J}(a,b) = \lambda_{LM}\Delta\mathcal{L}_{LM} - \lambda_{comp}\Delta\Phi + \lambda_{qual}\Delta Q + O(\epsilon^2)} \tag{4}$$

其中 $\epsilon = 1/|\mathcal{S}|$ 表示经语料库归一化的扰动。

证明。边际增益可依据标准词汇优化分析分解为三个组成部分（Sennrich 等人，2016年）。

语言模型分量：变化 $\Delta\mathcal{L}_{\text{LM}} \approx f(a,b) \cdot$点互信息$(a,b)$源自伪似然近似，其中点互信息（Pointwise Mutual Information）捕捉统计关联（Church 和 Hanks，1990）。
复杂度分量：根据最小描述长度原则（Rissanen，1978），合并会降低词汇复杂度：$\Delta\Phi = -\log|V| - 1 + O(|V|^{-1})$。这种压缩收益被纳入点互信息项，该项也倾向于高频共现。

质量分量：对于凹聚合函数 $g(x) = (x+\epsilon_Q)^\alpha$，詹森不等式给出 $g(\bar{q}_{ab}) \geq \frac{1}{2}(g(q_a)+g(q_b))$。主要的质量贡献为 $\Delta Q_+ = f(a,b) \cdot g(\bar{q}_{ab})$，其中 $\bar{q}_{ab} = (q_a + q_b)/2$。归一化误差为 $O(f(a,b)/T)$，对于典型语料库而言可以忽略不计。 □

C.3. 最优合并分数的推导

定理 C.3（质量感知合并评分——有理论依据的启发式方法）。受一阶近似 $\Delta\mathcal{J}$（引理 C.2）的启发，我们提出以下质量感知合并评分，作为一种有理论依据的启发式方法：

$$\boxed{w_{ab} = \frac{f(a,b)}{f(a)f(b)+\epsilon_f} \cdot (\bar{q}_{ab} + \epsilon_Q)^\alpha \cdot \psi(a,b)} \tag{5}$$

其中：

- $f(\cdot)$ 表示语料库中的频率
- $\bar{q}_{ab} = (q_a + q_b)/2$ 表示平均组成部分质量
- $\alpha \in (0, 1]$ 是一个用于控制质量敏感度的可学习参数
- $\epsilon_f, \epsilon_Q > 0$ 确保数值稳定性
- $\psi(a,b) \in [0,1]$ 编码了领域特定约束

注：以下推导涉及两项有原则的近似（步骤 4–5），以计算可处理性换取数学精确性。所得分数保留了关键的单调性，并通过下游任务性能进行端到端校准。

证明。由引理 C.2可知，边际增益为 $\Delta\mathcal{J}(a,b) = \lambda_{\text{LM}} f(a,b) \cdot$点互信息$(a,b) + \lambda_{\text{qual}} f(a,b) g(\bar{q}_{ab}) + O(1/|V|)$，其中复杂度项被吸收到点互信息中（二者都偏好频繁共现）。
逐次归一化：遵循 BPE 的设计原则（Sennrich 等人，2016年），我们按频率归一化，以捕捉逐次信息增益。应用指数变换（单调且保留排序）：$\exp(\Delta\mathcal{J}/f(a,b)) \propto f_{(a,b)}\, f(a) f(b) + \epsilon_f \cdot \exp(\frac{\lambda_{\text{qual}}}{\lambda_{\text{LM}}} g(\bar{q}_{ab}))$

幂律近似：我们用 $(\bar{q}_{ab}+\epsilon_Q)^{\tilde{\alpha}}$ 替换 $\exp(\lambda \cdot g(q))$，其中 $\tilde{\alpha}$ 通过端到端方式学习。这样既保留了 $\bar{q}_{ab}$ 的单调性，也涵盖了未知比率 $\lambda_{\text{qual}}/\lambda_{\text{LM}}$。最终分数为：$w_{ab} = f_{(a,\ b)}\, f(a) f(b)+\epsilon_f \cdot (\bar{q}_{ab}+\epsilon_Q)^{\alpha} \cdot \psi(a,\ b)$

单调性保证： $\partial w_{ab}/\partial \bar{q}_{ab} > 0$ 和 $\partial w_{ab}/\partial$基于点互信息的分数 > 0，确保优先选择质量更高且统计关联更强的合并。对 α 进行端到端学习，以校准启发式方法。 □

C.4. 推导中的关键见解

1. **点互信息基础：**频率项 $f_{(a,\ b)}\, f(a)\, f(b)+\epsilon_f$ 近似点互信息，用于捕捉统计关联。

2. **质量调节：**质量项 $(\bar{q}_{ab}+\epsilon_Q)^{\alpha}$ 以乘法方式调整基于点互信息的分数，提高高质量合并的权重。

3. **可学习的敏感度：**参数 α 控制质量与频率的相对重要性：
   - $\alpha = 0$：退化为标准的基于点互信息的词元化
   - $\alpha > 0$：提高质量信号的权重
   - 通过梯度下降学习，以优化下游性能

4. **领域灵活性：**因子 $\psi(a,\ b)$ 可在不修改核心框架的情况下融入领域知识。

这一推导表明，质量感知合并评分是一种有理论依据的启发式方法，其依据是对双层目标函数的第一性原理分析，而非对频率项和质量项进行临时组合。

C.5. 理论证明

定理的证明 3.2 （计算复杂度）。该双层优化问题可通过从加权集合覆盖进行多项式时间归约证明为 NP难（Karp，1972年）。该归约将集合映射为语料序列，将集合覆盖成本映射为词汇复杂度：给定一个加权集合覆盖实例 $(U, \mathcal{S}, \{c_i\})$，构造字母表 $\Sigma = U \cup \{\$\}$，为每个集合构造语料序列 σ_i，并赋予统一的质量分数。当 $\lambda_{\text{qual}} = 0$时，最优分词与最优集合覆盖一一对应。

关于证明一般双层规划 Σ_2^p-难的更强复杂度结果，参见（Bolte 等人，2024；Grne 与 Wulf，2023年；Dempe，2020）。最坏情况下，穷举搜索的复杂度为 $O(|\Sigma|^K \cdot K! \cdot N \cdot n \cdot |\Theta|)$，其中考虑了合并序列、合并顺序以及下游模型优化的搜索空间。

□

命题 C.4（有界性和利普希茨连续性 w_{ab}）。在假设 *(A1)-(A2)* 下，质量感知合并评分w_{ab} 在(q_a, q_b)中有界且利普希茨连续。

证明。**有界性：**根据 (A1)，$f(a,b)/(f(a)f(b)+\epsilon_f) \le C_f/\epsilon_f$。结合 $\bar{q}_{ab} \in [0,1]$和 $\psi \in [0,1]$，可得 $w_{ab} \le C_f(1+\epsilon_Q)^{\alpha}/\epsilon_f =: C_w$。

利普希茨连续性：根据紧致域上有界函数复合的链式法则，w_{ab} 关于 (q_a, q_b) 满足 L_w-利普希茨连续性，其中 $L_w = C_f L_g/\epsilon_f$。对于 $\alpha = 1$，$L_g = 1/\sqrt{2}$；对于 $0 < \alpha < 1$，$L_g \le \alpha \epsilon_Q^{\alpha-1}/\sqrt{2}$。正则化 ϵ_Q 确保数值稳定性。 □

命题 C.5（EMA 归一化的稳定性）。在假设 *(A1)* 和 $\epsilon_R > 0$ 下，对于非退化奖励流，基于 *EMA* 的归一化几乎必然保持 $\sigma_{j,t}^{\text{run}} > 0$。

证明。该结果可由标准 EMA 收敛理论（Robbins-Monro）得出。在 (A1) 条件下，原始奖励具有非退化分布 $\text{Var}(X_t) > 0$。EMA 方差更新可保持正性：若 $\text{Var}_{j,t-1}^{\text{run}} > 0$，则 $\text{Var}_{j,t}^{\text{run}} \ge (1-\beta_{\text{norm}})\text{Var}_{j,t-1}^{\text{run}} > 0$。

当 $\sum_t \beta_{\text{norm},t} = \infty$ 且 $\sum_t \beta_{\text{norm},t}^2 < \infty$ 时，运行方差几乎必然收敛到 $\text{Var}(X) > 0$，从而确保 $\sigma_{j,t}^{\text{run}} > 0$。 □

命题 C.6 (PPO 目标收敛性)。在假设 *(A1)-(A4)* 下，近端策略优化收敛到$J(\pi;\ \theta^{(0)}{}_{adapt})$的一个驻点。

证明。 在 (A1)–(A4) 下，标准的 PPO 条件成立（Schulman 等人，2017）：奖励有界（$|R(s,\ a)| \leq R_{\max}$）、状态空间紧、动作空间有限，且策略可微。裁剪替代目标确保梯度有界$\|\nabla_\theta L^{\text{CLIP}}\|_2 \leq G_{\max}$。

学习率 $\eta_t = \eta_0/\sqrt{t}$ 满足 $\sum_t \eta_t = \infty$ 且 $\sum_t \eta_t^2 < \infty$时，以$O(1/\sqrt{T})$ 的速率全局收敛至驻点，这可由（Bhandari 与 Russo，2021年；Cen 与 Chi，2023年）得出。 □

命题 C.7 （第二阶段梯度的一致性与有界性）。 在假设 *(A1)–(A3)* 下，*Gumbel-Softmax*估计器可产生一致梯度，且方差有界。

证明。 Gumbel-Softmax梯度的性质可由（Jang 等人，2017；Maddison 等人，2017）得出。在 (A1)–(A3) 下，复合逻辑值 ℓ_{ab} 有界于 $L_{\max} = C_w + \sum_j |\lambda_j| R_{\max}$。 Gumbel-Softmax雅可比矩阵满足$\|\partial y_i/\partial \ell_j\| \leq 1/\tau$，因此梯度 $\|\nabla_{\theta_{\text{adapt}}} L_{\text{task}}\| \leq L_{\max}/\tau \cdot \|\nabla_y L_{\text{task}}\|$有界。
当 $\tau \to 0$ 时，该估计量会收敛到 REINFORCE（Williams，1992）。偏差-方差权衡为：偏差$(\tau) = O(\tau^2)$，方差$(\tau) = O(1/\tau^2)$。最优温度 $\tau_{\text{opt}} \propto T^{-1/4}$针对 T个样本，对这两项进行权衡。 □

定理 C.8 （Gumbel-Softmax性质）。 设 $\pi = (\pi_1, \ldots, \pi_k)$ 是一个包含 k 个类别的分类分布。温度为 $\tau > 0$ 的 *Gumbel-Softmax* 分布满足：

1. 一致性： 当 $\tau \to 0$时，样本会收敛到分类分布中的独热向量(π)
2. 可微性： 重参数化可提供关于 π的连续梯度
3. 偏差-方差权衡：偏差 $O(\tau^2)$，方差 $O(1/\tau^2)$

证明。 这三项性质均已在（Jang 等人，2017；Maddison 等人，2017）中得到证明。我们总结其中的关键论证。
性质 1（一致性）： 根据 Gumbel-Max技巧，$\arg\max_i(\ell_i + g_i) \sim$ Categorical(Softmax(ℓ))，其中 $g_i \sim$ Gumbel(0，1)。当 $\tau \to 0$ 时，Gumbel-Softmax 样本 $y_i = \exp((\ell_i + g_i)/\tau)/\sum_j \exp((\ell_j + g_j)/\tau)$ 根据连续映射定理几乎必然集中于独热向量。

性质 2（可微性）： 对于 $\tau > 0$， y_i 在 ℓ_j 中是 C^∞，从而能够使用重参数化梯度。期望 $\mathbb{E}_g[y_i] =$Softmax(ℓ/τ)$_i$ 引入的偏差会在 $\tau \to 0$ 时消失。退火日程 $\tau_t \to 0$ 确保渐近一致性。
性质 3（梯度　界）： 雅可比矩阵满足 $\partial y_i/\partial \ell_j = (1/\tau) y_i(\delta_{ij} - y_j)$，从而得到 $\|\nabla_{\boldsymbol{\ell}} \mathbf{y}\|_F \leq 1/\quad \tau$. □

C.6. 假设

我们将理论分析中始终使用的假设形式化如下：

假设 A1（有界频率）： 存在 $C_f > 0$，使得对于所有词元 a， b:

$$0 \leq f(a), f(b), f(a,b) \leq C_f$$

假设 A2（有界质量）： 所有质量分数均满足 $q \in [0,1]$，且质量聚合函数是L_Q-利普希茨连续的。

假设 A3（有界奖励）： 原始奖励分量是有界的：对于所有 j， $|R_j^{\text{raw}}| \leq R_{\max}$.
假设 A4（正则学习率）： 学习率日程满足：- PPO： $\sum_t \eta_t = \infty$ 和 $\sum_t \eta_t^2 < \infty$ - 自适应学习： $\eta_t = O$（$1/\sqrt{t}$）

C.7. 理论扩展

定义 C.9 （近似保证的假设）。 $(1 - 1/e)$ 近似保证需要满足以下结构性假设：

（A1）**自适应单调性：** 对于任意部分实现 ψ 和合并 $(a,\ b)$： $\Delta_F((a,\ b)|\psi) \geq 0$，其中 Δ_F 表示边际增益。

（A2）**自适应次模性：** 对于实现 $\psi \preceq \psi'$（其中 $\preceq$ 表示扩展）： $\Delta_F((a,\ b)|\psi) \geq \Delta_F((a,\ b)|\psi')$（边际收益递减）。

(A3) **约束独立性：** $\psi(a,\ b)$ 与历史无关。(A4) **候选池正则性：** $\mathbb{P}[(a,\ b) \in PQ_t] \geq \delta > 0$ 对所有有效词对均成立。(A5) **质量稳定性：** $|q_t - \mathbb{E}[q_t|\mathcal{H}_t]| \leq \epsilon_q$ 以高概率成立。

引理 C.10 (近似自适应次模性)。 在假设 *(A3)-(A5)* 下，质量感知目标函数 $F(V) = \sum_k \mathcal{L}_{LM}(V; D_k) + \lambda_Q Q(V)$ 满足 ϵ-近似自适应次模性：

$$\Delta_F((a,b)|\psi) \geq \Delta_F((a,b)|\psi') - \epsilon_{sub} \tag{6}$$

对于 $\psi \preceq \psi'$，其中 $\epsilon_{sub} = O(\epsilon_q + 1/\delta)$。

证明概要。 基于频率的分量点互信息$(a,\ b)$ 具有精确的边际收益递减性质：随着合并次数增加，词对频率会降低，从而减少点互信息的潜在增益。质量分量$(\bar{q}_{ab})^\alpha$ 在 (A3) 下与历史无关，并在 (A5) 下保持稳定。近似误差 ϵ_{sub} 源于：(i) 质量估计噪声（ϵ_q），以及 (ii) 候选池变异性（$1/\delta$）。完整证明遵循Golovin 与 Krause (2011) 的框架。 □

定理 C.11 （具有显式常数的近似保证）。 在定义 *C.9*下，若假设 *(A1)-(A2)* 严格成立，则最大化 w_{ab} 的贪心策略可达到：

$$F(\pi_{greedy}) \geq \left(1 - \frac{1}{e}\right) F(\pi^*) - K\epsilon_q - \frac{K}{\delta}, \tag{7}$$

其中 π^* 是预算 K下的最优自适应策略。误差项来自 ϵ-近似次模性（引理 *C.10*）。

证明。 根据定理 5（出自 Golovin 与 Krause （2011）），对自适应次模函数进行贪心优化可达到 $(1-1/e)$ 近似。我们将此推广到 ϵ-近似次模性（引理 C.10）。

在 ϵ-近似次模性下，贪心每步保证变为 $\Delta_F((a_t,\ b_t)|\psi_t) \geq \frac{1}{K}[F(\pi^*) - F(\psi_t)] - \epsilon_{\text{sub}}$。定义 $\Delta_t = F(\pi^*) - F_t$并在 K步上迭代： $\Delta_K \leq (1-1/K)^K \Delta_0 + K\epsilon_{\text{sub}} \leq \frac{1}{e}F(\pi^*) + K\epsilon_{\text{sub}}$，利用 $(1-1/K)^K \leq 1/e$。

代入 $\epsilon_{\text{sub}} = \epsilon_q + 1/\delta$ 得到 $F(\pi_{\text{greedy}}) \geq (1-1/e)F(\pi^*) - K\epsilon_q - K/\delta$。 □

备注（假设与鲁棒性）：假设 (A1)–(A2)（自适应单调性和次模性）是 $(1-1/e)$ 保证的**充分条件**，但在实践中可能并不完全成立。具体而言：

- 语言模型损失 $\mathcal{L}_{\text{LM}}$ 通常不具有关于合并操作的次模性；该保证适用于所定义的质量-频率分量 $F(V)$。
- 当假设不成立时，界限将变为近似形式： $F(\pi_{\text{贪心}}) \geq (1-1/e)F(\pi^*) - K\epsilon_q - K/\delta - \epsilon_{\text{偏差}}$，其中 $\epsilon_{\text{偏差}}$与假设偏离程度成正比。

从经验来看，我们的实验表明该保证具有实际意义，原因如下：(1) 分词目标通常表现出近似次模行为（Lin & Bilmes，2011）；(2) α 端到端学习通过校准质量-频率权衡来弥补偏差；(3) 阶段 1 中强化学习策略的探索有助于摆脱纯贪心算法会陷入的局部最优解。

C.8. 鲁棒性分析

我们分析了在质量指标设定有误和质量分数具有对抗性的情况下的鲁棒性，并量化了强化学习与自适应学习阶段之间的交互效应。

定理 C.12 （对质量扰动的鲁棒性）。 设 $\tilde{q} = q + \xi$其中 $\xi \sim \mathcal{N}(0, \sigma_\xi^2)$。那么

$$\mathcal{L}(\tilde{q}) - \mathcal{L}(q) \leq \alpha\, \sigma_\xi \sqrt{\mathbb{E}[\|\nabla_q \mathcal{L}\|^2]}. \tag{8}$$

证明。 该结果由双层目标的利普希茨稳定性得出。根据链式法则和柯西–施瓦茨不等式，$|\mathcal{L}(\tilde{q}) - \mathcal{L}(q)| \le \|\xi\|_2 \cdot \int_0^1 \|\nabla_q \mathcal{L}(q + t\xi)\|_2 dt$。
根据命题 C.4，$\|\partial w_{ab}/\partial q\| \le \alpha \epsilon_Q^{\alpha-1}$，得到 $\|\nabla_q \mathcal{L}\| \le \alpha C_{\mathcal{L}}$。对 $\xi \sim \mathcal{N}(0, \sigma_\xi^2 I)$ 取$\mathbb{E}[\|\xi\|_2] \le \sigma_\xi \sqrt{d}$ 的期望，即可得到所述界。□

实证验证。 我们通过实验验证鲁棒性界：

- **20% 质量噪声**：性能分别下降 -4.2%（基因组学）和 -5.8%（金融），与 $O(\alpha\sigma_\xi)$ 界一致。
- **对抗性质量（分数反转）**：性能与标准字节对编码相当，符合质量信号变得无信息时的预期。
- **50% 质量缺失**：通过自适应 $\alpha \to 0$ 机制，平稳回退到仅基于频率的合并。

交互效应。 我们量化各学习阶段的贡献：

- 仅进行强化学习策略优化：相较于 BPE 的总改进幅度的 65%
- 仅进行自适应参数学习：总改进幅度的 45%
- 组合（协同效应）：联合优化额外带来 +10% 的提升

超加性效应（65% + 45% $>$ 100% 总计）表明这两个阶段会相互促进：强化学习会发现有前景的合并模式，随后由自适应学习进行校准；而学习得到的参数则会改善强化学习探索的奖励景观。

C.9. 信息论最优性

本小节证明，QA-Token 在噪声条件下达到信息论最优性，从而为质量感知分词提供理论依据。

命题 C.13（质量感知信息瓶颈解释）。 设 X 表示输入序列，T 表示分词表示，Y 表示下游任务标签。在具有质量分数 Q 的质量感知分词框架下，最优词汇表 V^* 最小化：

$$\mathcal{L}_{QA}(V) = -I(T;Y|Q) + \beta \cdot I(T;X|Q) \tag{9}$$

其中 $I(\cdot;\cdot|\cdot)$ 表示条件互信息，β 控制压缩-相关性权衡。

与合并分数的联系： 合并分数 w_{ab} 与（但并非直接推导自）该信息瓶颈目标相一致 。点互信息近似压缩效率 $I(T;X|Q)$，而质量加权确保在可靠区域中 $I(T;Y|Q)$ 较高。这种联系是定性的，而非通过对互信息进行直接微分得到，后者难以处理。

证明。 遵循信息瓶颈框架（Tishby 等人，1999）及其变分扩展（Alemi 等人，2017），以质量为条件 Q 得到目标函数 $\mathcal{L}_{\text{QA}} = -I(T;Y|Q) + \beta I(T;X|Q)$。
合并分数 $w_{ab} \propto$ 点互信息$(a,b) \cdot (\bar{q}_{ab})^\alpha$ 与此信息瓶颈目标函数一致：(i) 点互信息项近似于压缩效率 $I(T;X|Q)$（高点互信息的合并具有较高的压缩效率）；(ii) 质量项根据可靠性对合并进行加权，优先考虑 $I(T;Y|Q)$ 中的高质量区域。

注意：w_{ab} 的确切形式并非通过直接对互信息求导得出（这在计算上难以处理）。相反，它是一种有理论依据的启发式方法，通过端到端学习 α 来校准质量-压缩权衡。□

推论 C.14（降噪界）。对于噪声水平为 ϵ、质量分数为 q，且满足 $\mathbb{E}[q|\text{噪声}] < \mathbb{E}[q|\text{信号}]$, 的语料库，质量感知分词器可实现：

$$\mathcal{L}_{QA} \le \mathcal{L}_{uniform} - \alpha \cdot Var(q) \cdot \rho(q,\epsilon)^2 \tag{10}$$

其中 ρ（q、ϵ）表示质量分数与噪声水平之间的相关性。

关键理论见解。 这一信息论分析提供了三项基本见解：

1. **自动噪声过滤：** QA-Token隐式地执行重要性采样，在词汇表构建过程中提高高质量区域的权重。2. **最优压缩：** 质量感知合并过程通过将更多表示容量分配给高质量、信息丰富的区域，实现了更好的率失真权衡。

3. **迁移学习：** 使用 QA-Token 词汇表训练的基础模型能够学习到更稳健的表征，并更好地迁移至下游任务。

D. 完整的质量指标公式

D.1. 基因组学：详细的测序质量指标

在基因组测序中，每个核苷酸碱基判读 $s_i \in \{\text{A}, \text{C}, \text{G}, \text{T}, \text{N}\}$ 都对应一个 Phred 质量分数 $Q_{\text{phred},\ i} \in [0, 93]$:

$$P_{\text{error}}(i) = 10^{-Q_{\text{phred},\,i}/10} \tag{11}$$

碱基质量分数为 $q_i = 1 - P_{\text{error}}(i) \in [0, 1]$。位置调整质量考虑了读段末端的系统性质量下降:

$$q'_i = q_i \cdot \exp\left(-\beta_{\text{pos}} \cdot \frac{|i - (L-1)/2|}{(L-1)/2 + \epsilon_{\text{len}}}\right) \tag{12}$$

其中 L 为读长，$\beta_{\text{pos}} \geq 0$可学习，且 $\epsilon_{\text{len}} = 10^{-6}$。

对于多碱基词元 $t = s_1 ... s_{|t|}$，我们使用几何平均聚合:

$$q_t^{\text{genomic}} = \left(\prod_{j=1}^{|t|} q'_{s_j}\right)^{1/|t|} = \exp\left(\frac{1}{|t|}\sum_{j=1}^{|t|} \log(q'_{s_j} + \epsilon_Q)\right) \tag{13}$$

D.2. 金融：综合市场质量指标

金融时间序列质量综合了四个维度:

$$q_i^{\text{finance}} = \sum_{k=1}^{4} w_k \cdot q_{k,i}, \quad \sum_{k=1}^{4} w_k = 1 \tag{14}$$

1. 流动性质量：

$$q_{\text{liq}}(t) = \text{sigmoid}\left(\frac{\log(\text{volume}_t / \text{median_volume})}{\sigma_{\text{volume}}}\right) \tag{15}$$

其中 σ_{volume} 是在 $L_{\text{vol}} = 252$ 个交易日（一年）的回溯窗口上计算的对数成交量滚动标准差，并裁剪至 $[0.1, 10]$ 以确保数值稳定性。该归一化确保q_{liq} 相对于典型市场活动对成交量偏差作出成比例的响应。

2. 信号质量：

$$q_{\text{sig}}(t) = \max\left(0, 1 - \frac{|\text{bid-ask spread}_t|}{\text{mid-price}_t \cdot \alpha_{\text{spread}}}\right) \tag{16}$$

3. 稳定性质量：

$$q_{\text{stb}}(t) = \exp\left(-\beta_{\text{vol}} \cdot \frac{\text{realized_vol}_t}{\text{expected_vol}_t}\right) \tag{17}$$

其中预期_波动率$_t$ 是按照 RiskMetrics 方法计算的已实现波动率指数加权移动平均（J.P. Morgan，1996年）：预期_波动率$_t = \gamma_{\text{vol}} \cdot$ 预期_波动率$_{t-1} + (1 - \gamma_{\text{vol}}) \cdot$ 已实现_波动率$_{t-1}$，衰减因子为$\gamma_{\text{vol}} = 0.94$。可学习参数 $\beta_{\text{vol}} \geq 0$ 控制模型对波动率飙升的敏感度。

4. 信息质量：

$$q_{\text{info}}(t) = \frac{\text{MI}(\text{token}_t, \text{future_return}_{t+h})}{\text{H}(\text{future_return}_{t+h})} \tag{18}$$

词元聚合采用算术平均值：

$$q_t^{\text{finance}} = \frac{1}{|t|} \sum_{i \in t} q_i^{\text{finance}} \tag{19}$$

采用算术平均聚合的理由：与基因组学（使用几何平均数，公式13）不同，金融数据聚合出于两个有原则的理由采用算术平均值：（1）加性噪声模型：金融市场微观结构噪声在不同时间点上主要呈加性——单个含噪的跳动不会使相邻观测值失效，这与单个低质量 DNA 碱基会损害整条读段的情况不同。实证上，限价订单簿噪声源（延迟、部分成交、过时报价）的影响彼此独立，而非相乘。（2）用于预测的时间连续性：金融词元代表连续的时间窗口，下游任务（价格预测、波动率预测）基于窗口化特征运行。聚合质量自然体现了窗口内观测值的平均可靠性，这与预测模型对输入的加权方式相吻合。相比之下，基因组词元代表分子序列，其中任何不可靠碱基都会损害生物学解读（例如变异检测），因此需要采用较为保守的几何平均数，对哪怕单个低质量元素也予以惩罚。

D.3. 社交媒体：语言质量指标

社交媒体文本面临独特的质量挑战，包括拼写变体、语义漂移、平台特定惯例和时间动态。我们为字符级词元定义一个多维质量向量：

$$\mathbf{q}_t^{\text{social}} = (q_{\text{orth}}(t), q_{\text{sem}}(t), q_{\text{temp}}(t), q_{\text{plat}}(t)) \tag{20}$$

通过可学习的加权聚合得到标量质量：

$$q_t^{\text{social}} = \sum_j w_j \cdot q_j(t), \quad w_j \in \theta_{\text{adapt}} \tag{21}$$

1. 拼写质量：衡量与规范拼写的偏差：

$$q_{\text{orth}}(t) = \exp(-\lambda_{\text{edit}} \cdot d_{\text{edit}}(t, t_{\text{canonical}})) \tag{22}$$

其中，d_{edit} 是与参考词典中最近规范形式之间的归一化莱文斯坦距离。参考词典由以下内容组合构建：(i) Hunspell en_US 词典（2023年版，包含 ≈140k 个词条），(ii) 精选的社交媒体俚语词典（从 NoSlang.com 等来源汇总的 ≈50k 个词条），以及 (iii) 各基准测试任务的领域专用术语表。

2. 语义质量：捕捉上下文连贯性：

$$q_{\text{sem}}(t) = \max(0, \cos(\vec{v}_t, \vec{v}_{\text{context}})) \tag{23}$$

其中 $\vec{v}_{\text{context}}$ 是周围词元的平均嵌入。为提高效率，我们使用 fastText Common Crawl 嵌入（cc.en.300.bin，包含 2M 个词汇）（Bojanowski 等人，2017）。对于需要处理子词的 BERT 模型，我们使用`bert-base-uncased`，该模型来自 HuggingFace，并对所有子词词元进行平均池化。

3. 时间质量：模型对相关性随时间的衰减进行建模：

$$q_{\text{temp}}(t) = \exp(-\gamma_{\text{decay}} \cdot \Delta t) \tag{24}$$

with 发帖时间起的时间差，单位为 Δt 天，用于捕捉热门话题和时间相关性 e.

4. 平台质量：平台特定噪声建模：

$$q_{\text{plat}}(t) = P(t|\text{platform}) \tag{25}$$

使用 KenLM 训练的三元Kneser-Ney语言模型计算（Heafield，2011），训练数据为经过筛选的平台特定语料库（≈10每个平台各 M 个词元）：Twitter（包含 >100 个赞且含有 <5% 个特殊字符的推文）、Reddit（来自默认子版块且获得>10 个赞的评论）以及 Facebook（来自已验证页面的公开帖子）。这些“干净”子集为典型语言模式建立了平台特定基线。

学习得到的参数：对于 TweetEval 基准的实验，学习得到的参数为：$w_{\text{orth}} = 0.32 \pm 0.03$、$w_{\text{sem}} = 0.35 \pm 0.04$、$w_{\text{temp}} = 0.18 \pm 0.02$、$w_{\text{plat}} = 0.15 \pm 0.02$、$\lambda_{\text{edit}} = 0.5$ 和 $\gamma_{\text{decay}} = 0.01$。

E. 序贯学习过程：完整框架

本节详细介绍 QA-Token 两阶段序贯学习过程所采用的算法及收敛分析。

E.1. 阶段 1：强化学习策略优化

E.1.1. 马尔可夫决策过程形式化

词汇表构建过程被形式化为一个有限时域马尔可夫决策过程（完整说明请参见第 E.7 节）：

- **状态** $s_t \in \mathcal{S}$**：** 编码当前词汇表 V_t、合并候选项、语料库统计信息和进度 t/T
- **动作** $a_t \in \mathcal{A}_t$**：** 从优先队列中选择一对待合并项（a_i，b_i）
- **状态转移：** 合并操作之后确定性地更新词汇表
- **奖励：** 结合质量、信息和复杂度的多目标奖励

E.1.2. 奖励函数设计

奖励函数用于指导强化学习智能体：

$$R(a, b; \theta^{(0)}_{\text{adapt}}) = \sum_{j \in \{Q, I, C, \text{domain}\}} \lambda_j \hat{R}_j(a, b) \tag{26}$$

其中，各组成部分通过指数移动平均进行归一化（参见第 E.8 节）。具体组成部分如下：

- **质量奖励（**R_Q **来自** $R_{\text{原始}Q}$ **）：** 鼓励 $t_{\text{合并后}} = ab$具有较高的内在质量，使用领域特定聚合计算（第D节）。
- **信息奖励（**$\hat{}R_I$ **来自** $R_{\text{原始}I}$ **）：** 对统计显著的合并给予奖励，例如 $R_{\text{原始}I}(a, b) = \log P_{(t_{\text{合并后}})} P(a)P(b)+\epsilon_p$。
- **复杂度惩罚（**$\hat{}R_C$ **来自** $R_{\text{原始}C}$ **）：** 通常为负值，例如 $R_{\text{原始}C}(a, b) = -(|t_{\text{合并后}}| \cdot \log(|V_t| + 1))$. $\hat{}$ R_C 随后会缩放至例如 $[-1, 0]$。 $\hat{}$
- **领域特定奖励（**$R_{\text{领域 }k}$ **来自** $R_{\text{原始领域 }k}$**）：** 包括保守性分数（基因组学）和预测能力（金融）。

指数移动平均归一化奖励 $\hat{}$ $R_j(a, b)$ 由强化学习智能体在阶段 1 中使用。对于阶段 2 中的 Gumbel-Softmax logits（第 E.9 节），使用原始奖励分量或批归一化奖励组件，以确保相对于 θ_{adapt}的直接可微性。

E.1.3. PPO 训练算法

算法 1 阶段 1：强化学习策略训练

输入： 语料库 $\mathcal{S}$，初始 $\theta^{(0)}$ 适配，回合 E
初始化策略网络 π_{θ_π} 和价值网络 V_ϕ
对回合 $e = 1$ 至 E **执行**
  初始化词汇表 $V_0 = \Sigma$
  对于步骤 $t = 1$ 至 T **执行**
    根据当前词汇表计算状态特征 s_t
    采样动作 $a_t \sim \pi_{\theta_\pi}(a|s_t)$
    执行合并 $(a_{a_t},\ b_{a_t}) \mapsto ab$
    计算奖励 $r_t = R(a_{a_t},\ b_{a_t};\ \theta^{(0)}$ 自适应$)$
    存储轨迹 $(s_t,\ a_t,\ r_t)$
  结束循环
  使用 PPO 目标函数更新策略：
    $L^{\text{PPO}} = \mathbb{E}_t[\min(r_t(\theta)\hat{A}_t,\ \text{截断}(r_t(\theta)\ 1-\epsilon,\ 1+\epsilon)\hat{A}_t)]$
  更新价值网络以最小化均方误差
结束循环
输出： 优化策略 π_θ^* $_\pi$

E.2. 阶段 2：自适应参数学习

E.2.1. 自适应参数定义

可学习参数向量 $\theta_{\text{adapt}} \in \mathbb{R}^m$ 包括：

- **质量敏感度：** $\alpha \in [0, 2]$ 控制质量影响
- **领域因子：** $\beta_{\text{位置衰减}}$（基因组学位置衰减），$\beta_{\text{波动率}}$（金融波动率）
- **质量权重：** $\mathbf{w} = (w_1, \ldots, w_k)$，用于综合质量指标
- **奖励权重：** $\boldsymbol{\lambda} = (\lambda_Q, \lambda_I, \lambda_C, \ldots)$，用于多目标奖励

E.2.2. Gumbel-Softmax 可微优化

为了通过离散合并决策实现基于梯度的优化，我们采用 Gumbel-Softmax 松弛：

算法2 阶段 2：自适应参数学习

输入： 下游数据集 $\mathcal{D}$，策略 $\pi^*_{\theta_\pi}$，初始 θ_{adapt}
初始化温度 $\tau = \tau_{\text{init}}$
对于迭代 $i = 1$到 N **执行**
  从 $\mathcal{D}$中采样批次 B
  对于批次**中的每个**序列
    通过按 w_{ab} 排序的优先队列生成前 K 个合并候选项
    计算复合逻辑值：$\ell_{ab} = w_{ab}(a, b; \alpha) + \sum_j \lambda_j R_j^{\text{raw}}$
    使用 Gumbel-Softmax（可微松弛）选择合并：
      $y_i = \frac{\exp((\ell_i + g_i)/\tau)}{\sum_j \exp((\ell_j + g_j)/\tau)}$
    构建可微分词元化表示
  结束循环
  计算词元化批次上的任务损失 L_{task}
  更新参数：$\theta_{\text{adapt}} \leftarrow \theta_{\text{adapt}} - \eta \nabla L_{\text{total}}$
  退火温度：$\tau \leftarrow \tau \cdot \exp(-\beta_{\text{anneal}})$
结束循环
输出： 优化后的参数 θ^*_{adapt}

E.3. 最终词汇表构建

完成两个阶段后，便构建用于部署的最终词汇表。

详细流程：阶段 1（强化学习策略优化，得到 $\pi^*_{\theta_\pi}$）和阶段 2（自适应参数学习，得到 θ^*_{adapt}）完成后，通常会构建用于部署的最终词汇表。虽然存在多种可行策略，但我们的主要方法是利用优化后的自适应参数 θ^*_{adapt} 重新评估合并优先级。具体而言，从基础字母表开始执行贪心式 BPE 类过程。在每一步中，选择能使质量感知合并评分 $w_{ab}(a, b; \theta^*_{\text{adapt}})$ 最大化的合并操作 (a, b)；该评分根据公式 5 定义，并使用 θ^*_{adapt} 中的学习得到的参数（例如 α^*）。此过程持续进行，直至达到目标词汇表大小。或者，如果强化学习策略 $\pi^*_{\theta_\pi}$ 对于 θ_{adapt}的变化具有鲁棒性，也可以将其用于输入（状态特征、合并分数），这些输入使用 θ^*_{adapt}计算得出。不过，基于 $w_{ab}(\theta^*_{\text{adapt}})$的贪心方法通常更直接且计算高效，适用于部署，并利用了 θ^*_{adapt}所体现的对“优质”合并的精细理解。

算法 3 最终词汇表构建

输入： 语料库 $\mathcal{S}$，优化后的 θ^*_{adapt}，目标大小 K
初始化词汇表 $V = \Sigma$，合并次数 $m = 0$
当 $m < K$ **时执行**
  计算所有合并分数：$w_{ab} = \frac{f(a, b)}{f(a)f(b) + \epsilon_f} \cdot (\bar{q}_{ab} + \epsilon_Q)^{\alpha^*} \cdot \psi(a, b)$
  选择最佳合并项：$(a^*, b^*) = \arg\max_{(a,b)} w_{ab}$
  更新词汇表：$V \leftarrow V \cup \{a^* b^*\}$ // 组成部分 a^*，b^* 保留在 V
  更新语料库统计信息并重新计算受影响的频率
  $m \leftarrow m + 1$
9: </style>结束循环
输出： 最终词汇表 V^*

E.4. 收敛性质

序贯学习过程具有以下理论保证：

定理 E.1 (双时间尺度收敛)。 在假设 *A1-A4*（第 *C.6* 节）下，对 θ_π（快速时间尺度）和 θ_{adapt} （慢速时间尺度）的序贯优化以概率 *1* 收敛到局部纳什均衡。

证明。 该结果由双时间尺度随机逼近得出（Borkar，2009）。在 (A1)–(A4) 条件下，定理 2（Borkar，2009）的条件得到满足：(i) 利普希茨梯度（由有界奖励与光滑参数化得到），

(ii) 通过投影得到的有界迭代序列，(iii) 具有有界方差的鞅噪声，以及 (iv) 适当的步长（$\sum_t \eta_t = \infty, \sum_t \eta_t^2 < \infty$）。

在时间尺度分离 $\eta_\pi^{(t)}/\eta^{(t)}$ 适应 $\to \infty$ 的条件下，快速迭代 θ_π 会在 θ_{adapt} 发生显著移动之前达到均衡。迭代序列几乎必然收敛到满足 $\nabla_{\theta_\pi} J = 0$ 和 $\nabla_{\theta_{\text{adapt}}} L_{\text{total}} = 0$ 的极限点 $(\theta_\pi^*, \theta_{\text{adapt}}^*)$，从而构成局部纳什均衡。

□

关键性质：

- **阶段 1 收敛：** 近端策略优化收敛到驻点的速率为 $O(1/\sqrt{T})$（命题 C.6）
- **阶段 2 收敛 ：** Gumbel-Softmax 优化的收敛速率为 $O(1/\sqrt{T}) + O(\tau^2)$（命题 C. 7）
- **总体最优性：** 使用 θ^*_{adapt} 的贪心词汇表构建$(1 - 1/e)$可达到C.11 定理中的-近似

命题 E.2（具有显式常数的自适应学习收敛性）。在假设 ***A1–A4*** 下，给定 $\eta_t = \eta_0/\sqrt{t}$和 $\eta_0 \le 1/(2L)$，其中 L 是 ∇L_{total}的 *Lipschitz* 常数，我们有：

$$\mathbb{E}[\|\nabla L_{total}(\theta^T_{adapt})\|^2] \le \frac{2(L_{total}(\theta^0_{adapt}) - L^*)}{\eta_0\sqrt{T}} + \frac{4\eta_0 L\sigma^2}{\sqrt{T}}, \tag{27}$$

其中 L^* 是最优值，σ^2 界定梯度方差。

证明。该证 明遵循标准的非凸随机梯度下降分析（Kingma 和 Ba，2014）。根据 L_{total}的光滑性 ：

$$L_{\text{total}}(\theta^{t+1}) \le L_{\text{total}}(\theta^t) - \eta_t \langle \nabla L_{\text{total}}(\theta^t), g_t \rangle + \frac{L\eta_t^2}{2}\|g_t\|^2 \tag{28}$$

其中 g_t 是随机梯度。取期望并使用 $\mathbb{E}[g_t] = \nabla L_{\text{total}}(\theta^t)$ 和 $\mathbb{E}[\|g_t\|^2] \le \|\nabla L_{\text{total}}(\theta^t)\|^2 + \sigma^2$:

$$\mathbb{E}[L_{\text{total}}(\theta^{t+1})] \le \mathbb{E}[L_{\text{total}}(\theta^t)] - \eta_t(1 - L\eta_t/2)\mathbb{E}[\|\nabla L_{\text{total}}(\theta^t)\|^2] + \frac{L\eta_t^2\sigma^2}{2} \tag{29}$$

对 T次迭代进行望远镜求和，结合 $\eta_t = \eta_0/\sqrt{t}$ 和 $\eta_0 \le 1/(2L)$，即可得到所述界。

备注： 在温度退火 $\tau_t \to 0$下，Gumbel-Softmax 偏差项 $B(\tau)^2 \to 0$，从而确保渐近无偏。

□

定理 E.3 (局部与全局最优性)。 双时间尺度优化在局部强凸性条件下收敛至局部纳什均衡$(\theta_\pi^*, \theta_{adapt}^*)$，并满足质量界；概率重启可提高达到全局最优解的机会。

证明。 **第 1 部分：局部纳什均衡**。 根据定理 E.1，极限点满足 $\nabla_{\theta_\pi} J = 0$ 和 $\nabla_{\theta_{\text{adapt}}} L_{\text{total}} = 0$，构成局部纳什均衡。

第 2 部分：质量界。 在邻域 $\mathcal{B}_r(\theta^*_{\text{adapt}})$内， L_{total} 具有 μ-强凸性:

$$L_{\text{total}}(\theta^*_{\text{adapt}}) - L_{\text{total}}(\theta^{\text{global}}_{\text{adapt}}) \le \frac{1}{2\mu}\|\nabla L_{\text{total}}(\theta^*_{\text{adapt}})\|^2 = 0 \tag{30}$$

若全局最优解位于吸引域内。

第 3 部分：概率重启。 通过 M 次独立运行， $\mathbb{P}[\text{找到全局}] = 1 - (1 - p_{\text{basin}})^M \ge 1 - e^{-M \cdot p_{\text{basin}}}$，以 $\ge 1 - \delta$ 的概率在 $M = O\left(\log(1/\delta)/p_{\text{basin}}\right)$ 次重启内实现。 □

E.5. 算法概要

算法 4 QA-Token：质量感知分词框架

输入： 语料库 $\mathcal{C}$，质量分数 Q，词汇表预算 K
输出： 优化词汇表 V^*

阶段 1：强化学习策略优化
初始化策略 π_{θ_π}、自适应参数 $\theta^{(0)}$ 适应
对于回合 $e = 1$ 至 E **执行**
  $V \leftarrow \Sigma$(基础字母表)
  对于 步骤 $t = 1$ 至 K **执行**
    计算优先队列 PQ_t 及其分数 $w_{ab}(\cdot;\ \theta^{(0)}$ 适应)
    从 PQ_t中选择合并项 $(a,\ b) \sim \pi_{\theta_\pi}(\cdot|s_t)$
    执行合并： $V \leftarrow V \cup \{ab\}$ // 添加合并后的词元
    计算奖励 R_t 使用公式 26
  结束循环
  使用轨迹奖励进行近端策略优化更新 π_{θ_π}
结束循环

阶段 2：自适应参数学习
对于迭代 $i = 1$ 至 I **执行**
  采样合并候选项的小批量 $\mathcal{B}$
  使用公式 $\ell_{ab}(\theta_{\text{adapt}})$计算 logits，公式编号：37
  根据公式 38采样 Gumbel 噪声并计算软选择
  评估下游目标上的任务损失 L_{task}
  更新 $\theta_{\text{adapt}} \leftarrow \theta_{\text{adapt}} - \eta_i \nabla L_{\text{total}}$
结束 for

最终词汇表构建
使用贪心合并与 $w_{ab}(\cdot;\theta^*$ 构建最终词汇表 适配)
返回 V^*

算法 5 QA-Token 与下游 Transformer 的集成

1：输入：原始序列 X、训练好的 QA-Token 词汇表 V^*、Transformer 模型 M_θ ˆ 2：输出：任务预测 Y3： 4：// 分词（与 BPE 相比没有额外开销） 5： $T \leftarrow$ **Tokenize(**X, V^***) // 标准贪心分词 6： 7：// 嵌入和编码 8：** $E \leftarrow$ **TokenEmbed(**T**) +PosEmbed(位置) 9：对于层** $\ell = 1$**到** L**，执行 10：** $E \leftarrow$ **TransformerBlock**$_\ell$**(**E**) 11：结束循环 12： 13：// 任务头ˆ14：** $Y \leftarrow$ **TaskHead(**E**) // 分类、回归或生成 ˆ 15：返回** Y

算法 6 自适应参数的元学习初始化

要求：任务分布 $\mathcal{P}(\mathcal{T})$，基础初始化 $\theta^{(0)}$ 自适应，内循环步数 K，内循环学习率 η_{in}，外循环学习率 η_{out}1: 当未收敛时循环 2: 从 $\{\mathcal{T}_i\} \sim \mathcal{P}(\mathcal{T})$ 中采样任务批次 3: 对每个任务 $\mathcal{T}_i$ 执行 4: 设置 $\theta_i \leftarrow \theta^{(0)}$ 自适应 5: 对 $k = 1 \dots K$ 执行 6: 在 $\mathcal{T}_i$ 上计算 $L^{(i)}$ 总计(θ_i)，并更新 $\theta_i \leftarrow \theta_i - \eta_{\text{in}} \nabla_\theta L^{(i)}$ 总计(θ_i) 7: 结束循环 8: 结束循环 9: 更新初始化：$\theta^{(0)}$ 自适应 $\leftarrow \theta^{(0)}$ 自适应 $- \eta_{\text{out}} \sum_i \nabla_{\theta^{(0)}}$ 自适应$L^{(i)}$ 总计(θ_i) 10: 结束循环 11: 12: 返回元初始化 $\theta^{\star}_{\text{adapt}}$

E.6. 强化学习训练实现

E.6.1. 状态表示

在合并步骤 t，提供给强化学习智能体的状态 s_t 包括：

- **全局特征：** 当前词汇表大小为 $|V_t|$；剩余合并步数为 $T - t$；聚合词元统计信息（平均长度、质量分数的均值/标准差）。
- **候选对特征（优先队列中的前 K_{PQ} 个）：** 对于每个词元对（a、b）：频率 $f(a)$ $f(b)$ $f(a$、$b)$；质量 q_a、q_b；长度 $|a|$、$|b|$；合并分数 w_{ab}。
- **领域上下文：** 市场状态指标（金融）、平台标识（社交媒体）或序列质量（基因组学）。

近端策略优化智能体使用具有 2 个隐藏层（每层 256 个单元，ReLU 激活）的多层感知机。策略网络输出 K_{PQ} 候选项上的动作概率；价值网络输出单个标量。

E.6.2. 探索策略

采用 ϵ-贪心策略，并使用指数衰减在训练回合中将 ϵ 从 $\epsilon_0 = 0.5$ 退火至 $\epsilon_{\text{final}} = 0.05$，从而在全部领域中有效平衡探索与利用。

E.7. 马尔可夫决策过程形式化与细节

定义 E.4（分词化马尔可夫决策过程）。分词化马尔可夫决策过程是一个元组 $\mathcal{M} = (\mathcal{S},\ \mathcal{A},\ \mathcal{P},\ \mathcal{R},\ \gamma,\ T)$，其中：

1. **状态空间 $\mathcal{S}$：** 每个状态 $s_t \in \mathcal{S} \subset \mathbb{R}^d$ 编码：
   - 当前词汇表 V_t 及其统计信息（大小、词元长度分布）
   - 由排名靠前的合并候选项构成的优先队列 $PQ_t = \{\ (a_i, b_i, w_{a_i b_i})\ \}_{i=1}^{K_{PQ}}$
   - 语料库统计信息：频率分布、质量直方图
   - 进度指示器：t/T，其中 T 为合并预算

   形式上，$s_t = [\phi(V_t)\ \phi(PQ_t)\ \phi(\mathcal{S}_t)\ t/T] \in \mathbb{R}^d$。

   状态编码函数 ϕ： 编码函数 $\phi : \mathcal{X} \to \mathbb{R}^{d_x}$ 将可变大小的结构映射为固定维向量：

   - $\phi(V_t) = [|V_t|/|\Sigma|, \overline{|t|}, \sigma_{|t|}, \bar{q}_t, \sigma_{q_t}] \in \mathbb{R}^5$：词汇表大小比率、词元长度的均值/标准差、词元质量的均值/标准差。

- $\phi(PQ_t) \in \mathbb{R}^{6 \cdot K_{PQ}}$：对于前$K_{PQ}$个候选项，按每对拼接 $[w_{ab}, q_a, q_b, |a|, |b|, \log f(a,b)]$。若候选项数量不足，则用零填充。
- $\phi(\mathcal{S}_t) \in \mathbb{R}^{20}$: 质量直方图（ $[0,1]$上的$B_q = 10$ 个分箱）和对数频率直方图（观测范围内的$B_f = 10$ 个分箱）。

状态总维度： $d = 5 + 6 \cdot K_{PQ} + 20 + 1$。有 $K_{PQ} = 50$，我们有 $d = 326$。多层感知机策略网络通过两个隐藏层（256、128 个单元）和 ReLU 激活来处理这一表示（参见附录 E.6）。

2. **动作空间 $\mathcal{A}_t$：** 在时间 t:

$$\mathcal{A}_t = \{i : (a_i, b_i) \in PQ_t, i \le K_{PQ}\} \tag{31}$$

每个动作 $a_t \in \mathcal{A}_t$ 都会从优先队列中选取一对进行合并。

3. **转移动态 $\mathcal{P}$：** 确定性状态转移:

$$s_{t+1} = \mathcal{P}(s_t, a_t) = \text{UPDATE}(s_t, \text{MERGE}(a_{a_t}, b_{a_t})) \tag{32}$$

其中，MERGE 执行词汇表更新，而 UPDATE 重新计算统计数据。

4. **奖励函数：** $\mathcal{R}(s_t, a_t) = R(a_{a_t}, b_{a_t}; \theta^{(0)}$ 适应)

5. **折扣因子：** $\gamma = 1$(无折扣，有限时域)

6. **时域：** $T = K$ 次合并操作

命题 E.5 (马尔可夫决策过程良构性)。 分词化马尔可夫决策过程满足:

1. 马尔可夫性质： $P(s_{t+1}|s_t, a_t, s_{t-1}, \ldots\ldots) = P($
$s_{t+1}|s_t, a_t)$*2.* 有界状态空间： $\|s_t\|_2 \le C_s$*3.* 有限动作空间:
$|\mathcal{A}_t| \le K_{PQ}$

证明。(1) 源于状态包含状态转移的完整信息。(2) 成立，因为词汇表大小受$|\Sigma| + T$限制，且频率已归一化。(3) 则由优先队列的构造方式决定。 □

□

E.8. 奖励归一化细节

每个原始奖励分量 $R_j^{\text{raw}}(a, b)$ 都使用自适应运行统计量进行归一化。我们对均值 $\mu_{j,\,t}^{\text{run}}$ 和方差 $\text{Var}_{j\,t,}^{\text{run}}$ 维护指数移动平均（EMA）：

$$\mu_{j,t}^{\text{run}} = (1 - \beta_{\text{norm}})\mu_{j,t-1}^{\text{run}} + \beta_{\text{norm}} R_j^{\text{raw}}(a,b) \tag{33}$$

$$\text{Var}_{j,t}^{\text{run}} = (1 - \beta_{\text{norm}})\text{Var}_{j,t-1}^{\text{run}} + \beta_{\text{norm}}(R_j^{\text{raw}}(a,b) - \mu_{j,t-1}^{\text{run}})(R_j^{\text{raw}}(a,b) - \mu_{j,t}^{\text{run}}) \tag{34}$$

其中 $\beta_{\text{norm}} \in [10^{-3}, 10^{-2}]$。归一化分量为:

$$\hat{R}_j(a,b) = \frac{R_j^{\text{raw}}(a,b) - \mu_{j,t-1}^{\text{run}}}{\sigma_{j,t-1}^{\text{run}} + \epsilon_R} \tag{35}$$

使用 $\epsilon_R = 10^{-8}$ 以确保稳定性。

E.9. Gumbel-Softmax 梯度推导与温度退火

E.10. 温度退火日程

我们对温度参数采用指数退火调度：

$$\tau(t) = \tau_{\text{init}} \cdot \exp(-\beta_{\text{anneal}} \cdot t/T_{\text{anneal}}), \tag{36}$$

其中，$\tau_{\text{init}} = 1.0$、$\beta_{\text{anneal}} = 3.0$ 和 T_{anneal} 是优化步数总数。

该日程确保：

- **早期探索：** 较高的初始温度允许探索多样的合并模式
- **逐步细化：** 指数衰减使过渡到离散选择的过程更加平滑
- **收敛：** 较低的最终温度趋近于独热分类采样

E.11. 梯度计算

候选合并（a，b）的复合逻辑值为：

$$\ell_{ab}(\theta_{\text{adapt}}) = w_{ab}(a, b; \alpha) + \sum_j \lambda_j R_j^{\text{raw}}(a, b), \tag{37}$$

它们可通过合并分数和奖励权重相对于 θ_{adapt} 求导。

复合逻辑值结合了 w_{ab}（通过点互信息纳入频率，并通过 $\bar{q}_{ab}$ 纳入质量）与原始奖励分量R_j^{raw}，后者也体现了质量（R_Q^{raw}）和信息（R_I^{raw}）。

有意重叠的理由：尽管这些组件之间存在有意的重叠（两者都编码了质量和频率信号），但它们服务于不同的优化目的：

- w_{ab} **（合并分数）：**在阶段 1 通过强化学习目标（累积折扣奖励）进行优化，从而捕捉语料库级别的质量与频率权衡，并使其能够泛化到不同的合并序列。
- $\sum_j \lambda_j R_{\text{原始}j}$ **（加权奖励）：**在阶段 2 通过下游任务损失 $L_{\text{任务}}$进行优化，从而能够针对任务特定地重新加权质量、信息与复杂度。

这种分解使系统能够学习不同的质量与频率权衡，分别用于策略学习（阶段 1）和任务特定的适配（阶段 2）。参数 α 控制通过奖励最大化学得的通用词元质量偏好，而 λ_j 则根据任务特定梯度调整各项的相对重要性。消融研究（附录G.5，表13）证实，移除任一组件都会使下游性能下降 3–8%，验证了尽管两者存在重叠，其贡献仍缺一不可。

Gumbel-Softmax 分布提供了一种可微近似：

$$y_i = \frac{\exp((\ell_i + g_i)/\tau)}{\sum_{j=1}^{|\mathcal{C}|} \exp((\ell_j + g_j)/\tau)}, \quad g_i \sim \text{Gumbel}(0, 1) \tag{38}$$

任务损失的梯度通过蒙特卡洛采样计算：

$$\nabla_{\theta_{\text{adapt}}} L_{\text{task}} = \mathbb{E}_{\mathbf{g}} \left[\nabla_{\theta_{\text{adapt}}} L_{\text{task}}(\mathbf{y}(\boldsymbol{\ell}(\theta_{\text{adapt}}), \mathbf{g}, \tau)) \right] \tag{39}$$

其中$\mathbf{g}$ 是采样得到的 Gumbel 噪声。

梯度流： 梯度通过以下路径传播：

1. **任务损失：** L_{task} 评估下游数据上的性能 2. **软分词：** Gumbel-Softmax 提供可微的词元边界 3. **合并 logits：** ℓ_{ab} 取决于可学习的 θ_{adapt}4. **质量分数：**通过 α 和领域参数 β_{pos}，β_{vol}5. **奖励权重：**通过复合分数中的 $\boldsymbol{\lambda}$

F. 超参数敏感性分析

我们对 QA-Token 框架的关键参数进行了全面的敏感性分析：质量敏感度指数 α、主要质量奖励权重 λ_Q，以及金融领域特定波动率缩放指数 β_{vol}。针对每个参数，我们在指定范围内改变其取值，同时将其他所有超参数保持在最优值；这些最优值是在自适应学习阶段确定的。

结果汇总于表10，表明尽管在学习得到的参数值处性能最优，该框架对微小扰动并非过度敏感。当参数偏离其最优值时，性能平缓下降而非灾难性崩溃，这表明模型在超参数空间中占据了一个相当宽的吸引域。

表10. 超参数敏感性分析。报告了当关键超参数在其学习得到的最优值（以*标示）附近变化时，主要指标上的性能。数值为 $n = 5$ 次运行的均值。

参数	值	性能指标
基因组学 (QA-BPE-seq) - 指标：变体F1		
α (质量敏感度)	0.3	0.869
	0.5	0.879
	0.72*	**0.891**
	1.0	0.884
	1.5	0.872
λ_Q (质量奖励权重)	0.15	0.879
	0.25	0.886
	0.35*	**0.891**
	0.45	0.885
	0.55	0.878
金融 (QAT-QF) - 指标：夏普比率		
α (质量敏感度)	0.25	1.61
	0.50	1.68
	0.95*	**1.72**
	1.50	1.65
	2.00	1.58
β_{vol} (波动率缩放)	0.10	1.63
	0.30	1.69
	0.50*	**1.72**
	0.70	1.67
	1.00	1.60

G. 完整实验结果

本节全面呈现了所有领域的实验结果，包括详细分析、基础规模评估和消融研究。

G.1. 基因组学结果：详细分析

主要观察结果： QA-BPE-seq 在变异检测上的 F1 分数比 DNABERT-k 高 4.0 个百分点（0.891 对比 0.851），Hedges' $g = 8.2$ 显示效应量较大。与标准字节对编码（0.824）相比，提升了 6.7 个百分点。分类学分类的表现比领域标准的 k-mer 分词高 3.1 个百分点。序列重建性能提升了 16%，表明信息得到了保留。

核心见解：

1. **字节级模型会遭遇灾难性失败：** ByT5 和 CANINE 的推理速度慢 2.5×，准确率低 7-9%，这有力地证明了基于词汇表的分词对于基因组序列仍不可或缺。

2. **质量感知是可学习的：** 收敛后的参数（$\alpha = 0.72 \pm 0.03, \beta_{\text{pos}} = 0.014 \pm 0.002$）表明，最优质量敏感度可通过我们的自适应学习框架发现。

3. **改进机制：** 对生成的词汇表的分析表明，QA-BPE-seq 会创建与生物学单元（密码子、基序）对齐的词元，同时在易出错的连接处断开——这种行为在没有显式生物学监督的情况下自然出现。

G.2. 金融基础模型：详细结果分析

QAT-QF 在全部金融任务中表现出显著的一致性，零样本改进幅度从 7.3% 到 27.0

具体观察：

- 模型在市场体制检测（+11.6% F1 分数）和尾部风险估计（+18.0%）方面的优异表现表明，质量感知分词捕捉到了基于频率的方法所忽略的市场动态。
- 尤其值得注意的是，订单流失衡预测提升了 27.0%；这一任务对微观结构噪声高度敏感。

- 这些结果验证了我们的假设：在分词过程中引入质量信号，能够使基础模型学习到更稳健的金融时间序列表征。

G.3. 计算成本

训练时间。

- 标准字节对编码：5–10 分钟（5GB，CPU）
- QA-Token 阶段 1（强化学习）：30–36 GPU 小时（A100）
- QA-Token 阶段 2（自适应）：20–24 GPU 小时

内存需求。

- 优先队列：$O(K_{PQ} \cdot d)$（∼10MB，适用于 K_{PQ}=200）
- 质量统计：$O(|V| \cdot s)$（∼100MB，适用于 32K 词汇表）
- 词对频率：$O(|V|^2)$（∼4GB，适用于 32K 词汇表）
- 峰值：∼16GB 图形处理器

通过质量分层采样进行分层训练。 对于无法进行完整词汇表优化的大规模语料库，我们采用质量-方差重要性采样：数据点被抽样的概率与 $\text{Var}(q_i) + \epsilon_{\text{base}}$ 成正比，优先选择质量差异较大的区域，因为在这些区域中，分词决策的影响最大。

定义 G.1（分层训练记号）。 设 $\mathcal{C}$ 表示完整语料库，$\mathcal{S} \subseteq \mathcal{C}$ 表示一个子集，其 $|\mathcal{S}| = r \cdot |\mathcal{C}|$ 对于子集比例 $r \in (0, 1]$。定义：

- $\mathcal{L}(V;\ D) = \mathbb{E}_{x \sim D}[-\log\ P_{\text{LM}}(x|V)]$：在分布 D 上使用词汇表 V 时的预期语言建模损失
- $V_{\mathcal{S}}^* = \arg\min_V \mathcal{L}(V; \mathcal{S})$：子集 $\mathcal{S}$ 的最优词汇表
- $V_{\mathcal{C}}^* = \arg\min_V \mathcal{L}(V; \mathcal{C})$: 完整语料库的最优词汇表 $\mathcal{C}$

命题 G.2(分层训练界**).** 在以下假设下：

(A1) 损失 $\mathcal{L}(V;\cdot)$ 关于数据分布是 L-利普希茨的（对分布变化的敏感度有界 *t)*

(A2) 质量-方差重要性采样可达到有效样本量 $n_{eff} = r \cdot |\mathcal{C}|/(1+CV^2)$ 其中 CV 是重要性权重的变异系数

那么，在重要性采样子集上 $V_{\mathcal{S}}^*$ 学习到的词汇表满足：

$$\mathbb{E}[\mathcal{L}(V_{\mathcal{S}}^*;\mathcal{C})] \leq \mathcal{L}(V_{\mathcal{C}}^*;\mathcal{C}) + O\left(L \cdot \sqrt{\frac{1+CV^2}{r \cdot |\mathcal{C}|}}\right). \tag{40}$$

证明概要。 该界限可由标准重要性采样理论推出（Owen，2013年）。根据（A1），差值$|\mathcal{L}(V;\ \mathcal{S}) - \mathcal{L}(V;\mathcal{C})|$ 由 $\mathcal{S}$与 $\mathcal{C}$之间的分布差异控制。采用权重$w_i \propto \text{Var}(q_i) + \epsilon_{\text{base}}$ 进行重要性采样，会对高方差区域过采样，从而降低这种差异；而这些区域的分词质量最为重要。根据重要性采样的有效样本量公式，估计误差按 $O(1/\sqrt{n_{\text{eff}}})$ 缩放，因此得到所述界限。利普希茨假设（A1）确保在 $\mathcal{S}$ 上的优化可以迁移到 $\mathcal{C}$。 □

超大规模策略（>100TB）。

1. 质量分层采样（0.1–1%） 2. 分布式近端策略优化（8–32个图形处理器） 3. 对数据流进行带重放的在线强化学习 4. 内存映射频率表

成本效益。

- +5–30% 任务性能
- -15–20% 词元数量（推理更快）
- 一次性成本可分摊至多个应用

G.4. 基础模型规模结果（病原体检测，GUE）

表 *11.* 病原体检测基准（马修斯相关系数）。

Task	DNABERT-2	DNABERT-S	NT-2.5b-Multi NT-2.5b-1000g	METAGENE-1	METAGENE-1（QA-Token）
病原体检测（平均值）	87.92	87.02	82.43	79.02	92.96	**94.53**
病原体检测-1	86.73	85.43	83.80	77.52	92.14	**93.81**
病原体检测-2	86.90	85.23	83.53	80.38	90.91	**92.95**
病原体检测-3	88.30	89.01	82.48	79.83	93.70	**95.12**
Pathogen-Detect-4	89.77	88.41	79.91	78.37	95.10	**96.24**

G.5. 消融研究与补充实验

G.5.1. 强化学习算法消融

我们通过将近端策略优化替换为 GRPO 和 VAPO（实现遵循（Shao等人，2024年；Yue等人，2025））， 评估 QA-Token 对强化学习优化器选择的敏感性。在各个领域，下游性能保持稳定，词汇相似度也维持在较高水平（杰卡德 ≥ 0.95），证实了该框架的模块化特性。

表 *12.* 基因组理解评估（GUE）。除 COVID 使用 F1 分数外，全部评估指标均为马修斯相关系数。

Task	CNN	HyenaDNA	DNABERT	NT-2.5B-Multi	DNABERT-2	METAGENE-1	METAGENE-1（QA-Token）
TF-Mouse（平均值）	45.3	51.0	57.7	67.0	68.0	**71.4**	**72.8**
0	31.1	35.6	42.3	**63.3**	56.8	61.5	62.1
1	59.7	80.5	79.1	83.8	**84.8**	83.7	84.1
2	63.2	65.3	69.9	71.5	**79.3**	83.0	**84.5**
3	45.5	54.2	55.4	69.4	66.5	82.2	**83.3**
4	27.2	19.2	42.0	47.1	**52.7**	46.6	47.0
TF-HUMAN（平均值）	50.7	56.0	64.4	62.6	**70.1**	68.3	69.9
0	54.0	62.3	68.0	66.6	**72.0**	68.9	70.2
1	63.2	67.9	70.9	66.6	**76.1**	70.8	72.0
2	45.2	46.9	60.5	58.7	**66.5**	65.9	66.8
3	29.8	41.8	53.0	51.7	**58.5**	58.1	59.0
4	61.5	61.2	69.8	69.3	77.4	77.9	**78.5**
EMP（平均值）	37.6	44.9	49.5	58.1	56.0	66.0	**67.5**
H3	61.5	67.2	74.2	78.8	78.3	80.2	**81.0**
H3K14乙酰化	29.7	32.0	42.1	56.2	52.6	64.9	**66.0**
H3K36三甲基化	38.6	48.3	48.5	62.0	56.9	66.7	**67.8**
H3K4单甲基化	26.1	35.8	43.0	**55.3**	50.5	55.3	**56.1**
H3K4二甲基化	25.8	25.8	31.3	36.5	31.1	51.2	**52.3**
H3K4三甲基化	20.5	23.1	28.9	40.3	36.3	58.5	**59.5**
H3K79三甲基化	46.3	54.1	60.1	64.7	67.4	73.0	**74.1**
H3K9乙酰化	40.0	50.8	50.5	56.0	55.6	65.5	**66.5**
H4	62.3	73.7	78.3	81.7	80.7	82.7	**83.5**
H4AC	25.5	38.4	38.6	49.1	50.4	61.7	**62.8**
PD（平均值）	77.1	35.0	84.6	**88.1**	84.2	82.3	85.5
ALL	75.8	47.4	90.4	**91.0**	86.8	86.0	88.5
无TATA盒	85.1	52.2	93.6	94.0	94.3	93.7	**94.5**
TATA	70.3	5.3	69.8	**79.4**	71.6	67.4	73.5
CPD（平均值）	62.5	48.4	**73.0**	71.6	70.5	69.9	71.2
ALL	58.1	37.0	**70.9**	70.3	69.4	66.4	68.0
无TATA盒	60.1	35.4	69.8	**71.6**	68.0	68.3	69.5
TATA	69.3	72.9	**78.2**	73.0	74.2	75.1	76.3
SSD	76.8	72.7	84.1	89.3	85.0	87.8	**89.5**
COVID（F1 分数）	22.2	23.3	62.2	73.0	71.9	72.5	**73.3**
全球胜率	0.0	0.0	7.1	21.4	25.0	46.4	**57.1**

表 *13.* 不同强化学习算法的消融比较，包括训练时间（GPU 小时）、推理时间（毫秒/序列）以及相对于近端策略优化（PPO）的词汇表杰卡德相似度。

领域	配置（指标）	指标值	训练时间（GPU 小时）	推理（毫秒/序列）	词汇 Jaccard
基因组学	QA-Token（近端策略优化）	**0.891**	34.0	10.2	1.00
	QA-Token（GRPO）	0.890	32.5	10.3	0.98
	QA-Token（VAPO）	0.892	31.8	10.2	0.97
	QA-Token（DAPO）	0.889	34.2	10.4	0.98
金融	QA-Token（近端策略优化）	**1.72**	28.0	15.2	1.00
	QA-Token（GRPO）	1.71	26.5	15.3	0.96
	QA-Token（VAPO）	1.73	25.0	15.1	0.95
	QA-Token（DAPO）	1.70	28.5	15.2	0.96
社会	QA-Token（近端策略优化）	**74.5**	30.0	12.5	1.00
	QA-Token（GRPO）	74.2	29.0	12.6	0.97
	QA-Token（VAPO）	74.6	28.0	12.5	0.98
	QA-Token（DAPO）	74.3	31.0	12.7	0.97

G.6. 数据筛选基线：干净数据上的 BPE 与噪声数据上的 QA-Token

一个自然的问题是：仅将数据过滤为高质量数据并使用标准字节对编码，是否能够达到 QA-Token 的性能。我们通过在质量最高的前 20% 序列上训练 BPE（平均 Phred质量分数 ≥ 30），并将其与在完整噪声语料库上训练的 QA-Token 进行比较，评估这一数据整理基线。

表 *14.* 数据筛选基线比较（基因组学变异检测）。在噪声数据上使用 QA-Token 的表现优于在经过整理的干净数据上使用 BPE。

方法	训练数据	变体F1	p值
BPE（完整语料库）	100%（含噪声）	0.824 ± 0.004	< 0.001
BPE（排名前20%的干净数据）	20%（Q≥30）	0.847 ± 0.005	< 0.001
QA-Token	100%（噪声）	$\mathbf{0.891 \pm 0.004}$	—

主要发现：

- 数据整理（在干净数据上进行 BPE）优于在全部噪声数据上进行 BPE： $+2.8$% F1 分数（0.847 对比 0.824）。
- 然而，在全部噪声语料库 上，QA-Token 比在干净数据上的 BPE 高出 $+5.2$% F1 分数（0.891 对比 0.847， $p < 0.001$）.
- 这表明，与完全丢弃噪声数据相比，质量感知分词能从中提取更多价值。

G.7. 基因组学：真实世界数据集（ONT、UHGG）

数据集：(i) GIAB HG002 ONT 长读长数据（高错误率、第三代）；(ii) 统一人类肠道基因组（UHGG）集合（大规模、低错误率 NGS）。

结果：QA-BPE-seq 在两种市场状态下均持续优于基线。ONT（高错误率）结果：

表 *15*。ONT（GIAB HG002）结果。均值及 95% 置信区间，基于 $n = 10$ 次运行。

方法	变体F1	分类准确率 F1 分数	重建损失	推理时间（毫秒/序列）
标准字节对编码	0.795 ± 0.006	0.812 ± 0.007	0.388 ± 0.012	11.5 ± 0.3
SentencePiece	0.801 ± 0.005	0.825 ± 0.006	0.371 ± 0.011	11.6 ± 0.4
WordPiece	0.798 ± 0.006	0.819 ± 0.007	0.379 ± 0.013	11.5 ± 0.3
DNABERT-k (6-mer)	0.823 ± 0.004	0.846 ± 0.005	0.352 ± 0.010	11.2 ± 0.3
QA-BPE-seq (100%)	**0.864 ± 0.005**	**0.881 ± 0.004**	**0.305 ± 0.009**	**11.8 ± 0.4**
QA-BPE-seq（70%）	0.830 ± 0.005	0.845 ± 0.004	0.345 ± 0.009	11.9 ± 0.4
QA-BPE-seq（50%）	0.795 ± 0.006	0.810 ± 0.005	0.380 ± 0.010	12.0 ± 0.4
QA-BPE-seq（30%）	0.750 ± 0.006	0.760 ± 0.005	0.420 ± 0.011	12.1 ± 0.5

NGS（UHGG）结果:

G.8. 金融：高频股票（AAPL）

数据集与设置：AAPL 的 LOBSTER 高频限价订单簿数据。

结果：QAT-QF 可扩展至股票，在预测和交易评估指标方面优于基线 .

G.9. 金融：滚动窗口时间鲁棒性（BTC/USD，2023全年）

为展示超越单个季度的时间鲁棒性，我们采用严格的滚动窗口协议，将 BTC/USD 评估扩展至 2023年 全年的四个季度。每个季度的词汇表和下游模型仅使用该季度之前的数据进行训练。

表 *16.* UHGG（NGS）结果。 $n = 10$ 次运行的均值及 95% 置信区间。

方法	变体F1	分类单元准确率 F1 分数	重建损失	推理时间（毫秒/序列）
标准字节对编码	0.852 ± 0.003	0.881 ± 0.004	0.295 ± 0.008	9.8 ± 0.2
SentencePiece	0.860 ± 0.003	0.893 ± 0.004	0.280 ± 0.007	9.9 ± 0.2
WordPiece	0.855 ± 0.004	0.887 ± 0.005	0.286 ± 0.009	9.8 ± 0.3
DNABERT-k（6-mer）	0.875 ± 0.002	0.908 ± 0.003	0.264 ± 0.006	9.5 ± 0.2
QA-BPE-seq (100%)	**0.915 ± 0.003**	**0.935 ± 0.003**	**0.221 ± 0.005**	**10.1 ± 0.3**
QA-BPE-seq（70%）	0.878 ± 0.004	0.898 ± 0.004	0.250 ± 0.007	10.2 ± 0.3
QA-BPE-seq（50%）	0.842 ± 0.005	0.860 ± 0.005	0.276 ± 0.008	10.3 ± 0.3
QA-BPE-seq（30%）	0.790 ± 0.006	0.805 ± 0.006	0.310 ± 0.009	10.5 ± 0.4

表 *17.* AAPL 高频结果。均值及 $n = 10$运行次数的 95% 置信区间。

方法	收益预测 (%)	波动率均方根误差	状态准确率 (%)	Sharpe	推理时间（毫秒/序列）
标准字节对编码	63.1 ± 0.6	0.0125 ± 0.0004	75.8 ± 0.7	1.41 ± 0.06	14.8 ± 0.4
SAX	61.5 ± 0.7	0.0121 ± 0.0005	77.0 ± 0.6	1.38 ± 0.07	14.2 ± 0.3
BOSS	64.0 ± 0.5	0.0113 ± 0.0004	80.1 ± 0.5	1.53 ± 0.06	14.5 ± 0.4
QAT-QF	**69.8 ± 0.5**	**0.0085 ± 0.0003**	**87.9 ± 0.4**	**1.81 ± 0.08**	**15.0 ± 0.5**

表*18.* BTC/USD 在2023年的滚动窗口样本外夏普比率。每个季度均使用严格基于先前数据训练的模型。均值为 $n = 10$ 次运行的 95% 置信区间。

季度	QAT-QF Sharpe	BPE Sharpe	Δ (%)	市场背景
2023年第一季度	1.72 ± 0.07	1.32 ± 0.05	+30.3	复苏阶段
2023年第二季度	1.58 ± 0.09	1.21 ± 0.06	+30.6	盘整
2023年第三季度	1.45 ± 0.08	1.15 ± 0.07	+26.1	高波动性
2023年第四季度	1.68 ± 0.10	1.29 ± 0.06	+30.2	牛市
平均值	**1.61**	**1.24**	**+29.8**	—

关键观察： (i) QAT-QF 在全部市场状态下均保持稳定的提升（+26–31%）。(ii) 2023年第三季度波动率较高（出现相当于芝加哥期权交易所波动率指数的尖峰）；QAT-QF 的增益依然持续（+26.1%），展现出跨市场机制稳健性。(iii) 在市场条件各异的四个季度中表现一致，验证了模型的泛化能力并非局限于单一测试期。

H. 特定领域实例化

现在，我们将详细介绍 QA-Token 框架在三个不同领域中的实例化：基因组测序、社交媒体文本和量化金融。各领域的详细伪代码算法见第 H.9节。

H.1. 基因组学（QA-BPE-seq）

背景： 此实例化针对 DNA 或 RNA 测序读段的分析，这些读段常受碱基判读错误影响，可用于遗传变异检测、分类学分类或序列建模等应用。**原子元素与质量：** 基础字母表为 $\Sigma = \{\text{A、C、G、T/U、N}\}$。每个原子碱基的主要质量信息 s_i 来自 Phred 分数 $Q_{\text{phred},\ i}$。错误概率为 $P_{\text{error}}(i) = 10^{-Q_{\text{phred},\ i}/10}$，由此得到原子质量分数 $q_i = 1 - P_{\text{error}}(i)$。为对读段末端质量下降进行建模，对于长度为 L 的读段中位置为 i（从 0 开始计数）的碱基，其位置调整质量为：

$$q'_i = q_i \cdot \exp\left(-\beta_{\text{pos}} \cdot \frac{|i - (L-1)/2|}{(L-1)/2 + \epsilon_{len}}\right) \tag{41}$$

其中，$\beta_{\text{pos}} \geq 0$ 是 θ_{adapt}中的一个可学习参数。词元质量（q_t）：对于词元 $t = s_1 ... s_{|t|}$，我们使用经过位置调整的原子质量的几何平均数来计算其聚合标量质量：$q_t = (\prod_{j=1}^{|t|} q'_{s_j})^{1/|t|}$。几何平均数对低质量碱基较为敏感。在合并分数（公式 5）中，该 q_t 用于构成质量 q_a 和 q_b。合并分数（w_{ab}）：使用公式 5 计算分数，其中包括几何平均质量 q_a、q_b、可学习参数$\alpha \in \theta_{\text{adapt}}$以及 $\psi(a, b) = 1$。奖励分量（$R_{\textbf{genomic}}$）：总体奖励（公式 26）使用权重 $\lambda_j \in \theta_{\text{adapt}}$。具体的原始分量 R^{raw} 包括：

- $R_{\text{原始}Q}(a,\ b)$：新形成的词元 t_{ab}的质量。这是其几何平均质量：$R_{\text{原始}Q}(a,\ b) = q_{ab} = (\prod_{l=1}^{|a|+|b|} q'_{s_{ab'\ l}})^{1/(|a|+|b|)}$。
- $R_{\text{原始}I}(a,\ b)$：概率对数比：$R_{\text{原始}I}(a,\ b) = \log \frac{P_{(t_{ab})}}{P(a)P(b)+\epsilon_p}$。
- $R_{\text{原始}C}(a,\ b)$：复杂度惩罚：$R_{\text{原始}C}(a,\ b) = -|t_{ab}|$。
- $R_{\text{原始}bio}$(可选)：基于与已知基因组特征（例如基因、来自 dbSNP 等数据库的调控元件）的重叠度计算的领域特定奖励（Sherry 等人，2001年）。

原始分量使用自适应 EMA 方法进行归一化（公式 35）。**自适应参数（$\theta_{\textbf{adapt}}$）：** 包括 α、β_{pos}、奖励权重 λ_j，以及软频率/质量门控的潜在参数。**算法：** 采用两阶段学习过程（第 E 节）。优化强化学习策略（算法 1），然后通过优化下游任务目标来学习自适应参数 θ_{adapt}（算法2）。

H.2. 量化金融（QAT-QF）

背景： 该实例化侧重于分析含噪、非平稳的高频金融数据，用于价格走势预测或制定交易策略等任务。**原子元素与质量：** 原子元素 s_i 是来自高频数据的离散化事件（例如，限价订单簿事件的固定长度片段）。每个原子元素 s_i 被赋予一个标量质量分数$q_i = \sum_k w_k q_{k,\ i}$，其中$q_{k,\ i}$ 是归一化的质量分量（例如，q_{snr}，q_{liq}），而 w_k 是 θ_{adapt}中的可学习权重。

词元质量（q_t）：对于由原子元素 $\{s_i\}_{i \in t}$组成的词元 t，其聚合标量质量为算术平均值：$q_t = \frac{1}{|t|} \sum_{i \in t} q_i$。这用于 q_a，q_b，并应用于合并分数。合并分数（w_{ab}）：使用公式 5 计算，其中包含q_a、q_b、可学习的 $\alpha \in \theta_{\text{adapt}}$以及 $\psi(a,\ b) = 1$。市场状态：已识别的状态指标可用于调节强化学习策略和奖励分量。奖励分量（$R_{\textbf{finance}}$）：原始分量 R^{raw} 使用自适应 EMA 方法进行归一化。

- $R_Q^{\text{raw}}(a, b)$: Length-weighted average quality: $R_Q^{\text{raw}}(a, b) = \frac{|a|q_a + |b|q_b}{|a|+|b|}$.

• $R_{\text{原始}I}(a,\ b)$：跨市场状态混合的信息奖励：$R_{\text{原始}I}(a,\ b) = \gamma_{\text{市场状态}} \cdot I_{\text{正常}a}(b) + (1-\gamma_{\text{市场状态}}) \cdot I_{\text{压力}a}(b)$，其中 $I_{\text{市场状态}} = \text{对数}^{P}(t_{ab}|\text{市场状态})_{P}(a|\text{市场状态})_{P}(b|\text{市场状态})_{+\epsilon_p}$。混合因子$\gamma_{\text{市场状态}}$是 θ自适应中的可学习参数。

• $R_{\text{原始}P}(a,\ b)$：预测能力（与未来收益的互信息）：

$$R_P^{\text{raw}}(a,b) = \frac{\text{MI}(t_{ab}, \text{Disc}(R_\tau))}{\text{NormFactor}_{MI} + \epsilon_{MI}} \tag{42}$$

Disc(R_τ) 是离散化未来收益。NormFactor_{MI} 是自适应归一化因子。

• $R_{\text{原始}C}(a,\ b)$：采用波动率缩放的复杂度惩罚：

$$R_C^{\text{raw}}(a,b) = -\left(|t_{ab}| \cdot \log(|V_k|+1) \cdot \text{VolScale}\right) \tag{43}$$

其中，VolScale 取决于一个可学习参数 $\beta_{\text{vol}} \in \theta_{\text{adapt}}$。

自适应参数（θ_{adapt}）： 包括 α、质量分量权重 w_k, β_{vol}, γ_{regime}以及奖励权重 λ_j。**算法：**与基因组学领域一样，采用两阶段学习过程。

H.3. 社交媒体文本（QA-BPE-nlp）

背景：此实例旨在应对处理嘈杂的用户生成文本所面临的挑战，适用于情感分析或命名实体识别等任务。原子元素与质量：基础字母表由字符组成。词元 t 的质量由一个多维向量 $\boldsymbol{q}_t = (q_{\text{orth}}(t)\ q_{\text{sem}}(t)\ldots)$建模，详见附录 D.3。聚合标量质量为$q_t = \sum_j w_j \boldsymbol{q}_{t,j}$，其中 $w_j \geq 0$ 是 θ_{adapt}中的可学习权重。词元质量（q_t）：聚合分数 q_t 用于合并分数中的q_a, q_b。合并分数（w_{ab}）：使用公式 5 计算，其中包含 q_a, q_b、可学习的 $\alpha \in \theta_{\text{adapt}}$以及语义兼容因子 $\psi(a,b)$：

$$\psi(a,b) = \exp(\beta_{sem} \cdot \text{cosine}(\boldsymbol{v}_a, \boldsymbol{v}_b)) \tag{44}$$

其中 $\boldsymbol{v}_a$，$\boldsymbol{v}_b$ 是预训练嵌入，$\beta_{sem} \geq 0$ 是 θ_{adapt}中的可学习参数。**噪声模型：** 概率模型 $P(t'|t)$用于捕捉可能的变化，并为噪声鲁棒性奖励 R_N提供依据。**奖励分量（R_{social}）：** 原始分量经过归一化后，再由 λ_j加权。

• $R_{\text{原始}Q}(a,\ b)$：组合质量与直接质量的融合：$R_{\text{原始}Q}(a,\ b) = \omega^{\frac{|a|q_a+|b|q_b}{|a|+|b|}} + (1-\omega)q_{ab}$，可学习混合权重为 $\omega \in [0,1]$。

• $R_{\text{原始}S}(a,\ b)$：语义连贯性：点互信息$(a,\ b) \cdot$余弦_相似度$(\boldsymbol{v}_a,\ \boldsymbol{v}_b)$。

• $R_{\text{原始}N}(a,\ b)$：噪声鲁棒性：$R_{\text{噪声}}(t_{ab})$ _ $\frac{|a|R_{\text{噪声}}(a)+|b|R_{\text{噪声}}(b)}{|a|+|b|}$，基于噪声模型。

• $R_{\text{原始}C}(a,\ b)$：复杂度惩罚：$R_{\text{原始}C}(a,\ b) = -|t_{ab}|$。

• $R_{\text{原始}V}(a,\ b)$：词汇效率：$\log(1+f(t_{ab}))\,|t_{ab}|$。

自适应参数（θ_{adapt}）： 包括 α，β_{sem}、质量维度权重 w_j、奖励权重 λ_j以及混合权重 ω。**算法：**与其他领域一样，采用两阶段学习过程。

H.4. 金融实验方法详情

量化金融领域的全部交易模拟和收益预测评估（第5.2节）均严格遵循回测最佳实践，以确保结果的有效性并避免常见陷阱。

• **滚动前推验证：** 采用了严格的滚动前推验证方案。数据集被划分为按时间顺序排列的分段。对于每个分段 k，模型（包括 QA-Token词汇构建和下游预测/交易模型）均使用截至分段 k开始时的数据进行训练，在分段 $k-1$ 上进行验证（或使用训练数据中专门划分的验证部分），然后仅在分段 k上进行样本外测试。随后，训练窗口向前滚动，将分段 k 纳入训练，然后在分段 $k+1$上进行测试。此过程确保，对于特定测试期间，模型始终在训练或超参数调优阶段未见过的数据上接受测试。

• **前视偏差预防：** 我们采取了极为严格的措施，以防止任何形式的前视偏差。在任意时间 t，所有特征、质量分数、词元定义和交易决策均严格基于截至并包括时间 $t-1$时可获得的信息。用于训练预测模型或作为 R_P 奖励组成部分的未来收益标签（$R_{t+\tau}$）均取自严格晚于输入特征和词元构建所使用信息的时间段。

• **测试集和数据划分：** 整体数据集（2023年第一季度的 BTC/USD 限价订单簿数据）按时间顺序划分：70% 用作初始训练集，15% 用作验证集（用于下游模型的超参数调优和早停），最后 15%（约 2 周的 1 分钟级数据）用作最终样本外测试集，以报告夏普比率和预测准确率等最终评估指标。该测试集被留出，仅在所有模型开发和调优完成后使用一次。

• **交易成本：** 为模拟市场摩擦，每笔交易均按 5 个基点（0.05%）收取符合实际情况的交易成本。在交易模拟中，买入和卖出动作均会扣除该成本。

• **PPO 交易智能体详情：** 基于 PPO 的交易智能体使用一个 2 层多层感知机策略网络和一个独立的 2 层多层感知机价值网络，每个网络均包含 128 个隐藏单元，并采用 ReLU 激活函数。这些网络的输入由近期词元嵌入序列（由 QAT-QF 或基线分词器根据限价订单簿数据生成）以及智能体当前的市场头寸（做多、做空或空仓）组成。智能体的动作空间为离散型（买入、卖出、持有）。近端策略优化智能体的奖励函数是其在短时域内交易产生的已实现损益（PnL），并针对交易成本进行了调整。采用了标准的 PPO 超参数，包括裁剪参数 $\epsilon = 0.2$、广义优势估计 $\lambda = 0.95$ 以及用于探索的熵奖励。近端策略优化智能体在滚动前向方案中定期重新训练。

• **关于 $R_{\text{原始}P}$ 奖励（方程 42）：** 参数 M_{MI}（NormFactor 的窗口$_{MI}$）在我们的实验中设为 1000 个合并步数。未来收益 R_τ 对应 $\tau = 5$ 分钟前，并根据训练数据的经验分位数离散为 3 个区间（负向、中性、正向）。

H.5. 奖励分量详解

合并词元 a和 b为 $t_{merged} = a||b$时，奖励 $R(a,b)$ 的一般结构为： $R(a,b) = \sum_j \lambda_j \hat{} R_j(a,b)$，ˆ其中 R_j 为自适应归一化分量（参见第 4.2 节）。权重 $\lambda_j \geq 0$（通过 $\boldsymbol{\beta}_{\lambda_j}$ 和 Softmax 参数化）属于 θ_{adapt}。

H.6. 通用组件

• $R_{\text{原始}Q}(a,\ b)$：原始质量奖励。该分量旨在鼓励产生高质量词元的合并操作。原始分量的一种常见形式是组成词元 a和 b的质量的长度加权算术平均值：

$$R_Q^{\text{raw}}(a,b) = \frac{|a|q_a + |b|q_b}{|a| + |b|} \tag{45}$$

其中 q_a，q_b 分别是词元 a，b 的质量分数，$|a|$，$|b|$ 分别是它们的长度。对于社交媒体，$R_Q^{\text{raw}}(a,\ b)$ 可以采用混合方法：

$$R_Q^{\text{raw}}(a,b) = \omega \left(\frac{|a|Q_{agg}(a) + |b|Q_{agg}(b)}{|a| + |b|} \right) + (1-\omega) Q_{agg}(a||b) \tag{46}$$

其中 $Q_{agg}(t)$ 是词元 t (来自第 D.3节)的聚合质量分数，而 $\omega \in [0,1]$ 是 θ_{adapt}中的可学习混合权重。

• $R_{\text{原始}I}(a,\ b)$：原始信息增益。这会奖励统计显著的合并。一种常见的表述形式是：

$$R_I^{\text{raw}}(a,b) = \log \frac{f(t_{merged})}{f(a)f(b) + \epsilon_f} \tag{47}$$

其中 $f(\cdot)$ 表示频率，$\epsilon_f > 0$（例如 10^{-8}）用于稳定性。对于金融，可基于市场状态进行融合： $R_I^{\text{raw}}(a,b) = \gamma_{\text{regime}} I_{\text{normal}} + (1-\gamma_{\text{regime}}) I_{\text{stress}}$，其中 $I_{\text{regime}} = \log \frac{f(t_{merged}|M=\text{市场状态}) f(a|M=\text{市场状态}) f(b|M=\text{市场状态})}{+\epsilon_f}$。$\gamma_{\text{regime}} \in [0,1]$ 是 θ_{adapt} 中的可学习参数。

- $R_{\text{原始}C}(a, b)$：原始复杂度惩罚。该项对过于复杂的词汇表进行惩罚，且通常为负值。一种常见表述：

$$R_C^{\text{raw}}(a, b) = -\text{len}(t_{merged}) \cdot \log(|V_t| + 1) \cdot [\text{ScalingFactor}] \tag{48}$$

对于金融领域，缩放因子可以利用市场波动率，具体如下：$\beta_{vol} \in \theta_{adapt}$ 根据公式43。

H.7. 领域特定组件

• **基因组学：** $R_{\text{原始}bio}(a, b) = \text{分数}_{\text{重叠}}(t_{merged}$，已知生物学特征)。如果 t_{merged} 与已知生物学特征（例如，GENCODE 中的基因（Harrow 等人，2012年）、dbSNP 中的变异（Sherry 等人，2001年）显著重叠，则给予正奖励。重叠分数通过计算合并后的词元 t_{merged} 的字符跨度与已知基因组特征的字符跨度之间的杰卡德指数得出。杰卡德指数越高，表示重叠程度越大，获得的奖励也越高。

- **金融：**
  - $R_P^{\text{raw}}(a, b)$: 预测能力：

$$R_P^{\text{raw}}(a, b) = \frac{\text{MI}(t_{\text{merged}}; \text{Disc}(R_\tau))}{\text{NormFactor}_{MI} + \epsilon_{MI}} \tag{49}$$

使用互信息（MI）$\text{MI}(X; Y) = \sum_{x \in X,\ y \in Y} p(x, y) \log \frac{p_{(x,\ y)}}{p(x)p(y)}$。$R_\tau$ 是离散化未来收益（例如，根据训练数据的经验分位数将 $\tau = 5$ 分钟划分为 3 个区间）。$\overline{\text{NormFactor}}_{MI}$ 是根据当前强化学习回合中最近 M_{MI} 个合并步骤（例如，1000 步）内候选对的互信息值自适应计算出的第 95 百分位数。$\epsilon_{MI} > 0$（例如，10^{-8}）。尽管这种互信息自适应归一化会给强化学习回合中的 R_P 奖励成分带来一定程度的非平稳性，但研究发现，标准的近端策略优化训练能够妥善应对这一问题。奖励会根据新形成词元的信息量做出相应调整，这被认为是有益的；而 M_{MI} 窗口则能起到一定的平滑作用。使用固定归一化因子的替代方案（例如，根据互信息值的初始全局扫描结果计算）响应能力较弱，无法及时适应强化学习回合中词汇表演变时词元特征的变化。

- **社交媒体：**
  - $R_S^{\text{raw}}(a, b)$：语义连贯性：点互信息(a, b) ·余弦_相似度$(\boldsymbol{v}_a, \boldsymbol{v}_b)$。预训练嵌入 $\boldsymbol{v}_a, \boldsymbol{v}_b$（例如，fastText（Bojanowski 等人，2017））。– $R_N^{\text{raw}}(a, b)$：噪声鲁棒性：

$$\left(R_{\text{noise}}(t_{\text{merged}}) - \frac{|a| R_{\text{noise}}(a) + |b| R_{\text{noise}}(b)}{|a| + |b|} \right), \tag{50}$$

其中 $R_{noise}(t) = 1 - \mathbb{E}_{t' \sim P(\cdot|t)}[$归一化_编辑_距离$(t, t')]$ 基于噪声模型 $P(t'|t)$（附录 H.8）。– $R_V^{\text{raw}}(a, b)$：词汇效率：$\frac{\log(1+f(t_{\text{merged}}))}{|t_{\text{merged}}|}$。

H.8. 社交媒体噪声模型的更多细节

对社交媒体文本中的语言噪声进行形式化，需要定义从规范形式 $P(t'|t)$ 到观测变体 t 的概率变换 t' (Han 等人，2013年）。这些模型可用于计算噪声鲁棒性度量 $R_{\text{noise}}(t)$（定义见附录 H.5，公式 50）。$P(t'|t)$ 是根据社交媒体文本中常见错误模式总结出的启发式规则，以及现有噪声文本处理文献中阐述的原则构建的。建模的具体噪声类型包括：

- **字符级噪声：**

– **重复：** 字符 c 被实现为 c^n（一串 n 相同的字符）的概率。对于 $n \geq 1$，可以使用类几何分布对其建模。如果 p_{stop} 是不再额外重复一次的概率，则有：$P(c \to c^n) = (1 - p_{stop})^{n-1} \cdot p_{stop}$。参数 p_{stop} 经经验设定为 0.5，从而允许社交媒体中常见的适度重复（例如，“soooo goood”）。

– **替换：** $P(c_i \to c_j) = M_{\text{sub}}[c_i, c_j]$，其中 M_{sub} 是一个混淆矩阵。 M_{sub} 通过启发式方法构建，对标准QWERTY键盘布局上相邻字符之间的替换，以及常见的语音拼写错误（例如，'c' 与 'k'）赋予较高概率。非对角线概率通常较小。

– **遗漏（删除）：** $P(c \to \epsilon) = p_{\text{del}}(c)$ 是字符特定删除概率。该值被设为一个较小的均匀值（例如， $p_{\text{del}}(c) = 0.01$），适用于全部字符，以反映偶尔的意外遗漏。

- **词级噪声：**

– **缩写：** $P(w \to \text{abbr}(w)) = f_{\text{abbr}}(w \to \text{abbr}(w))$。该概率由一份汇编词典推导而来，该词典收录了源自公开可用在线语言资源的常见网络俚语和缩写。对于该词典中的词， f_{abbr} 被设为一个中等值（例如 0.3），否则为零。– **语音替换：** $P(w_1 \to w_2) \propto \exp(\lambda_{\text{phon}} \cdot \text{phon_sim}(w_1, w_2))$。语音相似度 $\text{phon_sim}(w_1, w_2)$ 使用 Double Metaphone算法计算得出。缩放因子 λ_{phon} 被设为 1.0。

- **话语级噪声（示例）：** 对于本文报告的实验，噪声建模主要聚焦于字符级和词级现象，因为这些现象非常普遍且易于建模。对话语级噪声进行显式建模，例如语码转换或复杂的标点模式，被认为超出了当前噪声组件 R_N 的范围，不过这代表了未来工作中一个有趣的研究方向。

这些概率模型用于定义 $P(t'|t)$，随后用于计算噪声鲁棒性度量 $R_{\text{noise}}(t) = 1 - \mathbb{E}_{t' \sim P(\cdot|t)}[\text{dist}_{\text{norm}}(t, t')]$ 中的期望距离。所使用的归一化距离度量 $\text{dist}_{\text{norm}}(t, t')$ 是莱文斯坦距离除以两个字符串 t 和 t' 的最大长度。

H.9. 特定领域算法

本节提供 QA-Token 框架在量化金融、基因组学和社交媒体中实例化的详细伪代码。这些算法补充了第 H 节中所述的领域实例化，展示了每个领域内的核心机制。

H.9.1. 量化金融（QAT-QF）

算法 7 质量感知分词合并评分与奖励计算（QAT-TOKEN - 金融）

要求： 当前词汇表 V_t，语料库统计信息（频率 $f(\cdot)$），当前自适应参数 $\theta_{adapt} = \{\alpha, \beta_{vol}, \gamma_{\text{regime}}, f_{min}, \delta_{\text{gate}}, w_k$（参数为 $\boldsymbol{\beta}_w$）$\}$，奖励权重 λ_Q，λ_I，λ_P，λ_C。
确保： 对于每个候选合并对（a，b）：质量感知合并评分 w_{ab}，即时总奖励 R（a，b）。
1：从语料库中识别候选合并对 C_t（例如，从优先队列 PQ_t 中识别）。
2：**对于**每个相邻词元对（a，b）$\in C_t$ **执行**
3：令 $t_{merged} \leftarrow a||b$。
4：检索/计算频率 f（a）、f（b）和 f（a，b）。
5：获取/计算平均质量 q_a，q_b（使用 $Q[i]$ 第D.2节中的数据，针对词元 a、b以及权重进行聚合）w_k =Softmax($\boldsymbol{\beta}_w$)$_k$）。
6：**质量感知合并评分**（w_{ab}）：$w_{ab} \leftarrow \frac{f(a,\ b)}{f(a)\cdot f(b)+\epsilon_f} \cdot \left(\left(\frac{q_a+q_b}{2}+\epsilon_Q\right)^{\alpha}\right) \cdot \psi$（$a$，$b$）// $\psi(a,\ b)=1$ 用于金融
7：**频率门控（可选）：** // 最终实验未使用频率门控 $\tilde{f}(a,b) \leftarrow f(a,b)$。
8：$R_Q^{\text{raw}}(a,\ b) \leftarrow \frac{|a|\cdot q_a+|b|\cdot q_b}{|a|+|b|}$.
9：根据市场状态条件估算 I_{normal}，I_{stress} $\tilde{f}(a,\ b)$。$R_I^{\text{raw}}(a,\ b) \leftarrow \gamma_{\text{regime}} \cdot I_{normal} + (1-\gamma_{\text{regime}}) \cdot I_{stress}$。
10：$MI_{val} \leftarrow$互信息(t_{merged}; 离散化(R_τ))。$R_P^{\text{raw}}(a,b) \leftarrow \frac{MI_{val}}{\text{归一化因子}_{MI}+\epsilon_{MI}}$ (归一化因子$_{MI}$ 来自第H.2节)。
11：$\sigma_{curr}, \sigma_{hist} \leftarrow$获取波动率(); $VolScaling \leftarrow (1+\max(0,(\sigma_{curr}-\sigma_{hist})/(\sigma_{hist}+\epsilon_{\text{vol}})))^{\beta_{vol}}$
12：$R_C^{\text{raw}}(a,b) \leftarrow -|t_{merged}| \cdot$对数($|V_t|+1$) $\cdot VolScaling$
13：归一化原始奖励：$\hat{R}_j(a,\ b) \leftarrow$自适应归一化($R_j^{\text{raw}}(a,\ b)$) 使用公式 35、33和 34。
14：**即时总奖励 (**$R(a,\ b)$**)：** $R(a,\ b) \leftarrow \sum_j \lambda_j$ $\hat{R}_j(a,\ b)$。
15：存储用于策略输入或选择的 (a，b) 的 w_{ab}, R(a，b) 及其他特征。
16：**结束循环**

算法 8 QA-TOKEN 的自适应参数学习（金融）

需要： 训练数据集 $\mathcal{D}_{\text{train}}$；下游任务损失函数 $L_{\text{task}}(\cdot,\ \cdot)$；模型参数 Θ_{model}；初始自适应参数 θ_{adapt}；学习率 η_θ；训练轮数 E_{adapt}；Gumbel-Softmax τ_g。
确保： 优化后的自适应参数 θ^*_{adapt}。
1：初始化 θ_{adapt}.
2：**对于**每个自适应训练轮次 $e=1,\ \ldots\ldots,\ E_{adapt}$ **执行**
3：**对于**每个小批量 $B=\{(S_{\text{seq},i}, Y_{\text{target},i})\}$ 从 $\mathcal{D}_{\text{train}}$ **执行**
4：$\mathcal{S}'_{batch} \leftarrow$ Gumbel 软令牌化($B, \theta_{adapt}, \tau_g$) // 等式37
5：$L_{\text{batch 任务_}} \leftarrow L_{\text{task}}(\mathcal{S}'_{batch}\ \{Y_{\text{target},i}\}, \Theta_{\text{model}})$
6：**如果**使用正则化 $L_{\text{reg}}(\theta_{adapt})$**则**
7：$L_{\text{total 批次_}} \leftarrow L_{\text{batch_task}} + L_{\text{reg}}(\theta_{adapt})$
8：**else**
9：$L_{\text{total_批次}} \leftarrow L_{\text{batch_task}}$
10：**结束如果**
11：计算梯度 ∇_θ $_{adapt} L_{\text{total_批次}}$. // 使用 Gumbel-Softmax 技巧
12：更新 $\theta_{adapt} \leftarrow \theta_{adapt} - \eta_\theta \nabla_{\theta_{adapt}} L_{\text{total_batch}}$。
13：将约束应用于 θ_{adapt}（例如 $\alpha \geq 0$、权重的 Softmax）。
14：**结束循环**
15：退火 τ $_g$.
16：**结束循环**
17：
18：**返回** $\theta^*_{adapt} \leftarrow \theta_{adapt}$。

H.9.2. 基因组学（QA-BPE-SEQ）

算法 9 合并奖励计算（基因组学）

要求： 词元 a， b 及其质量 q_a， q_b；频率 $f(\cdot)$；来自 θ_{adapt} 的奖励权重 λ_j。对于基因组学， q_a， q_b 表示组成词元的几何平均质量。
确保： 合并 a 与 b 的原始奖励 $R_j^{\text{raw}}(a, b)$。
$t_{merged} \leftarrow a||b$
$R_Q^{\text{raw}}(a, b) \leftarrow (\prod_{l=1}^{|t_{merged}|} q'_{s_{merged,l}})^{1/|t_{merged}|}$。 // 几何平均质量
$R_I^{\text{raw}}(a, b) \leftarrow$ 对数 $\frac{f_{(t_{merged})} \cdot f(a) \cdot f(b) + \epsilon_f}{}$。
$R_C^{\text{raw}}(a, b) \leftarrow -$长度$(t_{merged})$。
如果使用生物学奖励 **那么**
  $OverlapScore \leftarrow$ 计算重叠分数$(t_{merged}$，已知生物学特征$)$。
  $R_{bio}^{\text{raw}}(a, b) \leftarrow OverlapScore$。
结束如果

返回所有相关 $R_j^{\text{raw}}(a, b)$。（归一化奖励 $\hat{R}_j$ 稍后使用方程 35) 计算。

强化学习智能体的动作空间大小， K_{PQ} （即每个步骤中从优先队列选取的候选合并对数量），设为 $K_{PQ} = 50$。根据初步实验结果，该值能够在让强化学习智能体接触多种高潜力合并方案与保持动作空间规模适中（以实现高效的策略学习）之间取得良好平衡，因此选择了这一数值。对 $[20, 100]$ 范围内的数值进行探索后发现， $K_{PQ} \in [40, 60]$ 时性能相对稳定；较小的数值可能过早地剪除长期来看可能有益的合并方案，而较大的数值虽未带来显著收益，却会增加每个策略步骤的计算成本。最终选择的 50 能够在不同领域中有效平衡这些因素。

• 强化学习（PPO 具体设置）- 阶段 1：

– 策略/价值多层感知机架构：2-3 个隐藏层，每层包含 128-512 个单元。激活函数：ReLU 或 Tanh。– PPO ϵ_{clip} （裁剪参数）： $[0.1, 0.3]$，通常为 0.2。– GAE λ_{GAE} （广义优势估计 lambda）： $[0.9, 0.99]$，通常为 0.95。– 折扣因子 γ_{RL}： $[0.95, 1.0]$，对于不终止的任务或较长时域，通常为 0.99。– 优化器：Adam（Kingma 和 Ba，2014）。学习率 η_π （策略）、 η_v （价值）： $[1 \times 10^{-5}, 5 \times 10^{-4}]$。– 熵奖励系数 c_S （或 c_2）： $[0.0, 0.05]$，通常为 0.01。– 价值函数损失系数 c_{VF} （或 c_1）： $[0.25, 1.0]$，通常为 0.5。– 批大小（每次更新的状态转移数量）： $[128, 4096]$ 或更多，具体取决于数据/内存。– 每次更新的 PPO 更新轮数（对收集到的数据进行遍历的次数）： $[3, 20]$, , 通常为 $4 - 10$。– 行动者/并行环境数量：1 至 N_{cores} 或 N_{GPUs}。

• 自适应奖励归一化（第4.2节）：

– 指数移动平均动量 β_{norm}： $[10^{-3}, 10^{-1}]$，通常为 10^{-2}。– ϵ_R （稳定性常数）：通常为 10^{-8}。

• 奖励权重（β_{λ_j} 产生 λ_j）： β_{λ_j} 在 $\theta_{(0)\text{ 自适应}}$ 参数的初始值，在阶段 1 中可以为零或较小的随机数（从而得到均匀或近乎均匀的 λ_j）。随后会在阶段 2 中对其进行优化。

• 自适应学习参数（$\theta_{\text{自适应}}$ 来自算法 2）- 阶段 2：

– 优化器：Adam。学习率 $\eta_\theta \in [1 \times 10^{-6}, 1 \times 10^{-4}]$。– Gumbel-Softmax 温度 τ：从较高的初始值（例如 $1.0 - 5.0$）逐渐退火至较小的正值（例如 $0.1 - 0.5$），并在训练过程中完成。调度方式：例如指数衰减 $\tau_t = \max(\tau_{final}, \tau_0 \cdot d^t)$。– logit 复合函数（公式 37 ）： Norm_ℓ 通常为恒等变换；如果 logits 变化范围较大，则使用批归一化。

- **领域特定自适应参数和质量度量设置：**
  - **基因组学特定：**
    * β_{pos}（位置质量衰减）：学习得到。探索的初始范围为 $[0.001, 0.1]$。
    * ϵ_{len}（公式 41）：10^{-6}。
  - **社交媒体特定：**
    * $\boldsymbol{\beta}_{w_j}$（针对 Q_{agg} 权重 w_j）：通过学习得到。* β_{sem}（语义兼容性，公式 44）：通过学习得到。初始范围 $[0.1, 5.0]$。* ω（针对 R_Q^{raw}的混合权重，公式 46）：通过学习得到。通过无约束变量的 Sigmoid 函数进行参数化。* 注：在最终报告的社交媒体自然语言处理实验（第 H.3 节）中，未将直接下游损失分量 R_D 用于强化学习奖励。
  - **金融特定：**
    * $\boldsymbol{\beta}_{w_k}$ (适用于 $Q[i]$权重 w_k)：通过学习得到。* β_{vol}（适用于 R_C中的波动率缩放）：通过学习得到。初始范围为 $[0.0, 2.0]$。* γ_{regime}（适用于 R_I的状态混合）：通过学习得到。通过无约束变量的 Sigmoid 函数进行参数化。* M_{MI}（NormFactor 的窗口$_{MI}$）：例如，1000 步。* 注意：在量化金融实验的最终配置中，软频率门控已禁用（第5.2节）。

- **通用问答令牌参数：**
  – ϵ_f，ϵ_Q（方程5）：10^{-8}。– α（w_{ab}中的质量敏感度）：
  通过学习得到。初始范围为 $[0.0, 5.0]$。

- **词汇表设置：**
  – 目标词汇表大小 V_{target}：通常为 $[16000, 64000]$。

H.9.3. 收敛的自适应参数

表 19 给出了收敛均值（± 三次实验运行的标准差），对应于 θ_{adapt} 中每个领域的关键自适应参数。自适应学习过程会调节这些参数，以优化下游任务性能，从而形成特定领域配置。

表*19.* 收敛的自适应参数 (±标准差 D ev).

参数	基因组学	金融	社交媒体
α (质量敏感度)	0.72 ± 0.03	0.95 ± 0.03	1.15 ± 0.05
λ_Q (质量奖励权重)	0.35 ± 0.03	0.30 ± 0.02	0.33 ± 0.03
λ_I (信息奖励权重)	0.25 ± 0.02	0.20 ± 0.02	0.22 ± 0.02
λ_C (复杂度奖励权重)	0.15 ± 0.01	0.10 ± 0.01	0.12 ± 0.01
β_{pos} (基因组学位置衰减)	0.014 ± 0.002	N/A	N/A
β_{vol} (金融波动率缩放)	N/A	0.50 ± 0.05	N/A
γ_{regime} (金融状态融合)	N/A	0.60 ± 0.04	N/A
w_{orth} (自然语言处理正字法权重)	N/A	N/A	0.32 ± 0.03
w_{sem} (自然语言处理语义权重)	N/A	N/A	0.28 ± 0.02
w_{liq} (金融流动性权重)	N/A	0.45 ± 0.04	N/A
ω_{social} (自然语言处理质量融合)	N/A	N/A	0.55 ± 0.05

H.10. 社交媒体消融结果

表 20 中的消融研究旨在确认 QA-BPE-nlp 各质量感知组件的单独作用。我们区分以下影响：（1）多维质量奖励（行“无质量”），（2）语义连贯性考量（行“无语义”），（3）噪声鲁棒性特征（行“无噪声”），以及（4）自适应参数学习（行“无自适应参数”）。对所学权重 w_j 在各质量维度上的分析（具体数值详见附录 D.3）表明，不同维度的重要性有所差异（例如，正交性 q_{orth} 与语义 q_{sem} 在多次运行中往往获得更高权重），奖励分量 λ_i 也会随具体任务和数据集特征而自适应调整。

表20. 消融研究：QA-BPE-nlp 在 TweetEval情感分析上的结果。数值为均值及其 95% 置信区间，基于 $n = 10$ 次运行 s.

配置	TweetEval 分数	相对变化（%）
QA-BPE-nlp（完整）	**74.5 ± 0.3**	**-**
不含强化学习框架（贪心 w_{ab}）	72.1 ± 0.4	-3.2
不含质量（$R_Q = 0$）	71.5 ± 0.5	-4.0
不含语义（$R_S = 0$）	72.8 ± 0.3	-2.3
不含噪声（$R_N = 0$）	73.2 ± 0.4	-1.7
不含词汇效率（$R_V = 0$）	73.9 ± 0.3	-0.8
不使用自适应参数（α，w_j 固定）	71.8 ± 0.5	-3.6
QualTok-nlp（消融基线）	71.9 ± 0.4	-3.5

I. 数据集、基线方法和评估详情

本节补充介绍了正文讨论的数据集描述、基线方法和评估指标，并提供了理解和复现第5节所报告实验结果所需的更多细节。

I.1. 数据集与可复现评估

本小节详细介绍了第5节所报告实验中使用的具体数据集及其版本，以及相关的预处理步骤或配置。所有数据集均可公开获取，或依据学术研究许可证获取。

- **基因组学（QA-BPE-seq 实验）：**

  – 用于变异检测、重建和消融的模拟人类基因组读段：双端测序读段（150bp）以30倍覆盖度生成，所使用的是ART模拟器（版本2.5.8，采用`art_illumina` 工具）（Huang 等人，2012）。模拟基于GRCh38人类参考基因组（补丁13），并使用内置的HiSeq 2500错误谱（`-ss HS25`）。为严格评估高噪声场景下的稳健性，如第H.1节所述，该错误谱的默认碱基错误率（替换率与插入缺失率）相对于标准HiSeq 2500错误谱被人为加倍。关键ART参数包括：`-p -l 150 -f 30 -m 400 -s 10`。生成了约5GB的此类合成读段语料库，用于训练分词器、下游模型评估以及第H.1节所报告的消融研究。获取：ART模拟器为开源软件，可在以下地址获取 https://www.niehs.nih.gov/research/resources/software/art/。GRCh38参考基因组可从NCBI GenBank或Ensembl等公共存储库获取。
  – 用于变异检测评估的瓶中基因组（GIAB）真值集：变异检测性能以 HG002 真值集（v4.2.1，GRCh38）为基准进行评估（Zook 等人，2016年）。获取方式：GIAB 真值集可从 NIST FTP 站点公开获取。

  – 用于分类学分类的 CAMI II 宏基因组基准测试：分类准确率使用第二届CAMI挑战赛（Sczyrba 等人，2017）中的“玩具人类微生物组计划”（短读段，2019年8月组装）数据集进行评估。该基准测试提供了已知群落组成及相应测序读段的数据集，可用于性能评估。获取方式：CAMI II 数据集可通过 CAMI 挑战赛官方网站获取：https://data.cami-challenge.org/participate。

- **量化金融（QAT-QF 实验）：**

  – 加密货币限价订单簿（LOB）数据：BTC/USD 交易对的高频限价订单簿（LOB）数据取自学术数据服务 LOBSTER（https://lobsterdata.com/）（Huang & Polak，2011）。实验使用了 2023 年第一季度（Q1 2023）10 个档位的重建限价订单簿快照。如第 5.2 节所述，该数据集按时间顺序划分为 70% 的训练集、15% 的验证集和 15% 的样本外测试集。分词的原子元素定义为连续 5 个限价订单簿事件组成的序列，特征化方式见附录H.2。获取方式：LOBSTER 公开提供样例数据，完整数据集则需依据学术或商业许可证获取。

- **社交媒体文本（QA-BPE-nlp 实验）：**

  – TweetEval 基准： TweetEval 基准（Barbieri 等人，2020）用于在多种推文分类任务上评估 QA-BPE-nlp。TweetEval 为七项各不相同的任务提供了一个统一框架，采用标准化的数据集划分（训练集、验证集、测试集）和评估指标，具体包括：
  * 情绪识别（SemEval-2018 任务 1（Mohammad 等人，2018））* 表情符号预测（SemEval-2018 任务 2（Barbieri 等人，2018））* 反讽检测（SemEval-2018 任务 3（Van Hee 等人，2018））* 仇恨言论检测（SemEval-2019 任务 5（Basile 等人，2019））* 冒犯性语言识别（SemEval-2019 任务 6（Zampieri 等人，2019））* 情感分析（SemEval-2017 任务 4（Rosenthal 等人，2017））* 立场检测（SemEval-2016 任务 6（Mohammad 等人，2016））

  如第 I.8 节所述，实验包括使用不同的分词策略，在这些任务上微调预训练 BERTweet-base 模型（Nguyen 等人，2020）。访问方式：TweetEval 基准测试，包括

每个组成数据集的数据访问脚本和详细信息均可在 GitHub 上获取：
https://github.com/cardiffnlp/tweeteval。访问底层推文内容通常需要通过推文 ID 获取内容，并遵守 Twitter 的服务条款及相应的数据集许可证。

I.2. 数据集与发布计划

为使此前无法使用的嘈杂语料库能够用于基础模型训练，我们将发布：

- **分词器产物：**各领域（基因组学、金融、社交媒体）在多种词汇表大小下的最终 QA-Token 词汇表、合并表以及 θ适配。
- **基础模型就绪语料清单：**用于重建大规模嘈杂预训练语料库的脚本与清单（包括过滤和去重），以及与我们的 20 亿子集分词器训练方案相匹配的采样器配置。
- **评估套件：**面向基因组学（变异检测、宏基因组学）、金融（预测、波动率、市场状态、交易）和社交媒体（TweetEval）的可复现流程，以及强化学习消融测试框架。
- **文档与治理：**许可证、数据使用注意事项，以及高影响应用（例如财务决策和临床基因组学）中的负责任使用指南。

所有代码和制品　都将以宽松的学术许可证发布，以最大限度地提高可复现性和采用率　　n.

I.3. QA-Foundation：嘈杂预训练语料库提案

我们提出 QA-Foundation，这是一个精选套件，包含规模极其庞大的嘈杂语料库，专为支持基础模型规模的预训练而设计，并配有明确的质量标注和治理机制：

- 基因组学：涵盖数拍字节级的宏基因组测序读段（SRA），并附有规范化元数据、Phred质量分布、重复序列图谱、污染标记和逐条读段的来源哈希。质量数据通道包括逐碱基Phred质量值、平台、运行信息、修剪日志和接头污染。
- 金融：包含股票、期货和加密货币的多资产高频限价订单簿数据流，配有同步日历、微观结构指标（价差、深度、订单失衡）、市场状态标签以及交易所特定异常标记。
- 社交媒体/网络文本：来自多个平台的用户生成文本，带有时间戳、平台标签、去标识化的稳定作者哈希、规范化注释（话题标签、提及、URL）以及噪声变换（变体簇、重复、键盘距离混淆矩阵）。

每个领域都提供标准化模式、质量通道和采样清单，以便在多个规模（例如 0.1%、1%、5%）下复现分词器训练，并支持公平比较。脚本会生成清单、去重索引（MinHash/LSH）和质量审计报告。治理措施包括明确的许可证、预期用途声明和红队风险评估。我们将发布：

- 可供分词器使用的分片，附有校验和及完整性清单
- 质量通道提取器（开源）和验证套件
- 可复现采样器，可匹配我们针对基因组学的 2B 基础子集协议，以及其他领域的相应预算

I.4. 基线方法

我们实现并配置了以下基线分词方法，以便与第5节中提出的 QA-Token 变体进行严格比较。

- **标准字节对编码（BPE）**（Sennrich 等人，2016年）：传统的基于频率的合并算法。在基因组学和社交媒体实验中，除非另有说明，均使用 HuggingFace ‘tokenizers’ 库（0.15.0版）实现，并专门配置为 $tokenizers.models.BPE$（$unk_token =$”$[UNK]$” $min_frequency = 2$）。在量化金融实验中，则使用了可比的标准字节对编码实现。

• **SentencePiece** （Kudo 和 Richardson，2018）：一种无监督文本分词器和反分词器。在基因组学和社交媒体实验中，使用了 SentencePiece（0.1.99版）的字节级 BPE 模式，直接对原始文本进行处理。

• **WordPiece** (Wu等人，2016年)：一种因广泛应用于 BERT 而闻名的子词分词算法。它通过迭代合并词对来构建词汇表，使训练数据在一元语言模型假设下的似然最大化。

• **DNABERT k-mer** (Ji 等人，2021年)：在基因组学领域的实验中，采用固定 k-mer 分词作为强基线，具体使用 6-mer。这与 DNABERT 等模型中的常见做法一致。

• **符号聚合近似 (SAX)** (Lin 等人，2003)：一种成熟的时间序列数据符号表示方法，应用于量化金融实验。中间价序列采用窗口大小为 16、字母表大小为 8 的分段聚合近似 (PAA) 进行离散化。

• **SFA 符号袋 (BOSS)** (Schäfer，2015年)：一种时间序列分类算法，使用符号傅里叶近似 (SFA) 生成符号词（词元）。该方法作为量化金融领域的基线，应用于中间价序列。

• **QualTok（消融基线）**：如第 5 节所述，QualTok 是 QA-Token 的消融基线。它采用简化的质量感知合并评分，$w_{ab} \propto f_{(a,b)}^{\ } \cdot \left(\frac{q_a+q_b}{2}+\epsilon_Q\right)^{\alpha}$，但关键在于省略了针对合并序列的强化学习策略优化，以及用于复杂 θ 自适应参数的完整自适应学习循环（调优 α 之外的参数）。合并操作通常基于此分数采用贪心方式执行。

对于所有基线方法，我们依据文献中的常见做法（Sennrich 等人，2016年；Kudo 和 Richardson，2018；Wu等人，2016年；Devlin等人，2019；Brown等人，2020；Ji 等人，2021年）、这些方法原始实现中的具体建议，或通过对合理取值进行系统性扫描，从留出的验证集中找出表现最佳的配置，以确保比较稳健，来选择所有基线方法的关键超参数，例如目标词汇表大小（通常对应预先设定的合并操作次数，例如第 5 节按领域指定的 16,000 或 32,000 次）。

I.5. 评估指标

QA-T的性能至所呈现的结果 oken 和基线方法通过以下特定领域的评估指标进行评估，这些指标对应在第5节中。 d-

- **基因组学：**

  – **变异检测：**通过与 GIAB 真值集对比，使用 F1 分数、精确率和召回率评估性能。这些评估指标使用“hap.py”工具（版本 0.3.14）计算，该工具可从以下网址获取：`https://github.com/Illumina/hap.py`。– **分类学分类（宏基因组学）：**对于 CAMI II 基准测试，主要使用分类准确率评估性能（具体而言，使用表1中报告的整体分类性能 F1 分数）。– **序列重建损失：**还通过训练 Transformer 自编码器模型，并测量其在留出测试集上的重建损失（例如，离散词元的交叉熵），来评估词元表示的质量。

  变异检测模型架构：变异检测评估使用 Transformer编码器，将词元嵌入作为输入特征。该模型按位置输出变异概率（单核苷酸变异、插入、删除、参考序列）。训练时使用针对 GIAB HG002标签的交叉熵损失。此方法评估分词在学习得到的表征中保留变异信息序列特征的能力，并使用 hap.py 基准测试工具（v0.3.14）进行评估。

- **量化金融：**

  – **收益预测准确率：**对未来中间价收益率（例如，提前 5 分钟）的方向预测正确的百分比。– **波动率预测均方根误差：**预测的 5 分钟波动率与已实现波动率（根据更高频率的数据计算）之间的均方根误差。– **市场状态识别准确率：**将各时间段分类为离散市场状态（例如，由 GARCH-HMM 识别的两种状态）时达到的准确率。– **交易表现：**主要指标是在词元化数据上运行的基于PPO的交易智能体所达到的年化夏普比率（Sharpe，1994）。每笔交易计入 5 个基点的交易成本。我们还监测了最大回撤（MDD）和卡玛比率等其他表现指标（详情见附录H.4）。

- **社交媒体文本：**

  – 七项 TweetEval 基准任务的表现，按照基准测试组织者为各项任务指定的官方评估指标进行衡量（Barbieri 等人，2020）。这些指标包括：
  * 表情符号预测：准确率（Acc） * 情绪识别：宏平均 F1 分数（F1 M） * 仇恨言论检测：宏平均 F1 分数（F1 M） * 反讽检测：准确率（Acc） * 冒犯性语言识别：宏平均 F1 分数（F1 M） * 情感分析：宏平均召回率（Rec M） * 立场检测：各主题的平均 F1 分数（F1 Avg）

第 5 节中报告的所有实验结果均表示 $n = 10$ 次独立运行的均值和 95% 置信区间，以确保结果稳健，并便于评估变异性。

I.6. 代码可用性

我们将在 GitHub 上以 MIT 许可证发布 QA-Token 框架。该代码库包含源代码、配置文件、预训练模型，以及所有实验的可复现性脚本。

I.7. 近似 QA-Token：迈向计算高效的质量感知

QA-Token 的学习框架由于采用强化学习和自适应学习阶段，计算成本较高。未来工作将探索计算成本更低的近似方法。一个起点是我们的消融基线 QualTok，它采用基于质量感知分数的贪心合并策略 w_{ab}（公式 5），无需显式进行强化学习策略优化，从而避开阶段 1 强化学习的成本。

还可以通过以下方式进一步降低成本：

1. **精简贪心合并的自适应参数学习：** 我们无需进行完整的强化学习，而是可以专注于自适应地学习一组经过改进的参数 θ^*_{adapt}（例如，α、质量权重 w_j、简化后的奖励权重 λ_j），直接优化由贪心 w_{ab}引导、面向下游任务的分词。这既保留了质量感知适应能力的核心，又显著降低了复杂度，无需学习强化学习策略。基于 Gumbel-Softmax 的学习（阶段 2）将针对这些贪心合并优化 θ_{adapt}，也可以使用简化的复合逻辑值。

2. **策略蒸馏：** 如果强化学习策略 $\pi^*_{\theta_\pi}$ 能够捕捉复杂的合并依赖关系，则可以减轻部署时的计算开销。可以通过策略蒸馏（Hinton 等人，2015年；Rusu 等人，2016年）训练一个紧凑的“学生”模型（例如，较小的神经网络或决策树），使其模仿较大且经过预训练的“教师”强化学习智能体的决策，从而加快词汇表构建。

3. **代理辅助自适应学习：** 对 θ_{adapt}（阶段 2）的优化可以通过使用评估成本更低的代理模型（Jones 等人，1998）来近似下游任务损失 L_{task}，从而减少使用完整下游模型进行频繁且成本高昂的端到端评估的需求。

4. **迁移与元学习用于 θ_{adapt}：** 将从某个任务或数据集中学到的 θ_{adapt} 参数用作其他任务或数据集的初始化（如算法中所示Algorithm 6）可显著减轻新应用的训练负担。

I.8. 扩展的 TweetEval基准测试方法

本节描述 es 遵循全面的 TweetEval基准测试方法。结果见表 21 .

数据集与评估框架： TweetEval（Barbieri 等人，2020年）提供了一个统一框架，用于评估模型在七项异构推文分类任务上的表现，每项任务都具有固定的训练集、验证集和测试集划分。这使得不同方法之间能够进行标准化比较。这七项任务包括：情绪识别（Mohammad 等人，2018）（4个标签：愤怒、喜悦、悲伤、乐观）、表情符号预测（Barbieri 等人，2018）（20个表情符号标签）、反讽检测（Van Hee 等人，2018）（2个标签：反讽、非反讽）、仇恨言论检测（Basile 等人，2019）（2个标签：仇恨、非仇恨）、冒犯性语言识别（Zampieri 等人，2019）（2个标签：冒犯、非冒犯）、情感分析（Rosenthal 等人，2017）（3个标签：积极、中性、消极），以及立场检测（Mohammad 等人，2016）（3个标签：支持、中立、反对，涵盖五个主题）。对于每项任务，我们均使用 TweetEval 基准测试所规定的统一评估指标报告性能。表21提供了基线比较框架。文中报告了 TweetEval 所定义的每项任务的官方指标（详情另见 *https://github.com/cardiffnlp/tweeteval* ）。

表*21.* TweetEval基线比较框架。

模型	表情	情感	Hate	反讽	冒犯	情感	立场	全部（TE）
BERTweet	33.4	79.3	**56.4**	**82.1**	79.5	73.4	71.2	**67.9**
TimeLMs-2021	**34.0**	**80.2**	55.1	64.5	**82.2**	**73.7**	**72.9**	66.2
重新训练的 RoBERTa	31.4	78.5	52.3	61.7	80.5	72.8	69.3	65.2
基础版 RoBERTa	30.9	76.1	46.6	59.7	79.5	71.3	68.0	61.3
Twitter 版 RoBERTa	29.3	72.0	49.9	65.4	77.1	69.1	66.7	61.4
FastText	25.8	65.2	50.6	63.1	73.4	62.9	65.4	58.1
LSTM	24.7	66.0	52.6	62.8	71.7	58.3	59.4	56.5
SVM	29.3	64.7	36.7	61.7	52.3	62.9	67.3	53.5
QA-BPE-nlp + BERTweet	**34.2**	**81.5**	**58.8**	**82.9**	**83.0**	**75.1**	**73.5**	**70.0**