将阅读理解形式化为“认知资本”的交易与增值过程：基于传统数学的严格论证

- 作者：GaoZheng

- 日期：2025-09-20
- 版本：v1.0.0

摘要

本文围绕：首先明确问题背景与约束，给出可验证的形式化定义与工程接口；随后分解系统/模型/数据/指标的关键设计，并给出可复现的实现与对齐路径；最后总结风险与边界条件，给出落地建议与扩展路线。

定义 1（信息资产与分片）
给定有限序列空间

$$\mathcal{D}=\{C_1,\dots,C_T\},\quad T\in\mathbb{N},$$

其中每个片段 C_t 属于某一基空间 $\mathcal{X}$（例如词元序列的有限集）。记 $\mathcal{D}_{\text{chunk}}=\{C_1,\dots,C_T\}$。

定义 2（认知资本空间与代数）
认知资本空间 $(\mathbb{S},\oplus,\mathbf{0})$ 是幂等、交换、结合的幺半群（即**可交换幂等幺半群**）：

$$x\oplus y=y\oplus x,\quad (x\oplus y)\oplus z=x\oplus(y\oplus z),\quad x\oplus x=x,\quad \mathbf{0}\oplus x=x.$$

$\mathbb{S}$ 的元素表征“已提炼的结构化知识”，$\oplus$ 表示**无重复聚合**（幂等性对应“避免重复进仓”）。

定义 3（状态、动作、成本）
令上下文预算 $B_t\in\mathbb{N}$（可理解为 token/检索/验证预算）。
设状态空间

$$\mathbf{S}=\mathbb{S}\times\mathcal{D}_{\text{chunk}}\times\mathbb{N},\quad s_t=(\mathcal{S}_{t-1},C_t,B_t).$$

动作空间 $\mathbf{A}$ 为一组**可组合的基本算子**（如
ACQUIRE/EXTRACT/LINK/VERIFY/HEDGE/TRIM/COMMIT），形式化为可测映射族

$$a:\ \mathbb{S}\times\mathcal{D}_{\text{chunk}}\times\mathbb{N}\to\mathbb{S}\times\mathbb{N},\quad a(s_t)=(\Delta\mathcal{S}_t,\Delta B_t),$$

并据此更新

$$\mathcal{S}_t = \mathcal{S}_{t-1} \oplus \Delta\mathcal{S}_t, \qquad B_{t+1} = B_t - \mathrm{cost}(a, s_t) + \mathrm{refund}(a, s_t).$$

定义**交易成本** $c : \mathbf{S} \times \mathbf{A} \to \mathbb{R}_{\geq 0}$，例如 token/时间/验证滑点。

定义 4（MDP 与确定性转移）
令 MDP $\mathcal{M} = (\mathbf{S}, \mathbf{A}, P, R, \gamma)$，其中

$$P(s_{t+1} \mid s_t, a_t) = \delta\big((\mathcal{S}_t, C_{t+1}, B_{t+1})\big),$$

即转移由序列索引递进与 $\oplus$ 更新确定给出；R 为终止回报（见定义 6），$\gamma \in (0, 1]$ 为折扣（有限时域可取 $\gamma = 1$）。

定义 5（原子事实与特征抽取）
存在可测映射 $\mathcal{F} : \mathbb{S} \to 2^{\Omega}$ 将资本映为**原子事实集合**；Ω 为有限基底（主题/事实/关系标签）。

定义 6（估值函数与终值回报）
给出单调次模函数 $g : 2^{\Omega} \to \mathbb{R}_{\geq 0}$（定义见命题 1），并定义终值

$$\mathcal{V}(\mathcal{S}_T) = g\big(\mathcal{F}(\mathcal{S}_T)\big).$$

回报：

$$R(s_t, a_t) = \begin{cases} 0, & t < T, \\ \mathcal{V}(\mathcal{S}_T) - \lambda \sum_{k=1}^{T} c(s_k, a_k), & t = T, \end{cases}$$

$\lambda \geq 0$ 为成本权重。

二、马尔可夫性与可解性

引理 1（马尔可夫性）
在状态定义 $s_t = (\mathcal{S}_{t-1}, C_t, B_t)$ 与确定性转移下，过程 $\{s_t\}_{t=1}^{T}$ 关于策略 $\pi(a \mid s)$ 是马尔可夫的。

证明：由于 $\mathcal{S}_t$ 由 $(\mathcal{S}_{t-1}, C_t, B_t, a_t)$ 通过 $\oplus$ 与可测更新唯一决定，且文档序列索引确定推进 C_{t+1}，故 $\mathbb{P}(s_{t+1} \mid s_{1:t}, a_{1:t}) = \mathbb{P}(s_{t+1} \mid s_t, a_t) = \delta\big((\mathcal{S}_t, C_{t+1}, B_{t+1})\big)$。■

定理 1（最大熵目标的良定性）
定义最大熵目标

$$J(\pi) = \mathbb{E}_{\tau \sim \pi}\left[\mathcal{V}(\mathcal{S}_T) - \lambda \sum_{t=1}^{T} c(s_t, a_t) + \alpha \sum_{t=1}^{T} \mathcal{H}\big(\pi(\cdot \mid s_t)\big)\right]$$

在有限时域、有限/可数动作集下良定，且存在最优策略 $\pi^{\star}$。

证明：有限时域下，期望为有界和；$\mathcal{H} \ge 0$，$\mathcal{V} \ge 0$，成本非负，目标下界存在；策略空间在逐点拓扑下紧（有限情形）或可取劣化极限；标准极值存在性（Weierstrass 型）给出存在解。■

三、潜能塑形与策略不变性

定义 7（潜能函数与塑形）
给定 $\Phi : \mathbb{S} \to \mathbb{R}$，令增广奖励

$$r_t' = r_t + \gamma\Phi(\mathcal{S}_t) - \Phi(\mathcal{S}_{t-1}).$$

定理 2（潜能塑形不改变最优策略）
在固定 $\gamma \in (0,1]$ 与有限时域下，以 r_t' 替换 r_t 的 MDP 与原 MDP 具有**同一最优策略集**。

证明：沿轨迹望向和望远镜恒等式：

$$\sum_{t=1}^{T} \big(\gamma\Phi(\mathcal{S}_t) - \Phi(\mathcal{S}_{t-1})\big) = -\Phi(\mathcal{S}_0) + \gamma^T\Phi(\mathcal{S}_T).$$

这是与策略选择无关的常数（$\mathcal{S}_0$ 固定，$\Phi(\mathcal{S}_T)$ 在终值已被 $\mathcal{V}$ 吸收或可合并为等价常量项），不改变任意两策略回报差，故最优策略不变。■

四、估值的次模结构与近似最优保证

定义 8（单调与次模）
对函数 $g : 2^\Omega \to \mathbb{R}$，若 $A \subseteq B \Rightarrow g(A) \le g(B)$ 则单调；若对任意 $A \subseteq B \subseteq \Omega$ 与 $x \notin B$，

$$g(A \cup \{x\}) - g(A) \ \ge\ g(B \cup \{x\}) - g(B),$$

则称 g 次模（边际收益递减）。

命题 1（覆盖-多样-冗余惩罚的次模性）
令

$$g(F) = \underbrace{\sum_{u\in U} w_u \cdot \mathbf{1}\{\exists f \in F:\ f\text{覆盖}u\}}_{\text{覆盖}} + \underbrace{\sum_k \psi_k\big(\#\{f \in F:\ f \in \text{类}k\}\big)}_{\text{多样}} - \underbrace{\sum_{(f,f')\in F^2} \rho(f,f')}_{\text{冗余惩罚}},$$

其中 ψ_k 凸、非减、$\rho \ge 0$ 满足双线性小惩罚，则 g 单调次模（在 ρ 充分小的条件下保持次模，或采用已知次模冗余项如 facility-location 形式）。■

定理 3（预算下贪心的 $1-1/e$ 近似）
设每步新增原子集合的成本可加 $\mathrm{cost}(f)\ge 0$，总预算 B。令最优集合为 $F^\star$。经典贪心（每步选取单位成本边际增益最大的原子/片段）得到 F^{gr} 满足

$$g(F^{\mathrm{gr}}) \ \ge\ (1-1/e)\, g(F^\star).$$

证明要点：标准次模最大化在 knapsack 预算下的近似保证（可用比例贪心或连续贪心 + 多项式时间舍入）。证明基于边际收益递减与指数型衰减界，详见次模最大化的经典推导（此处给出结论与核心不等式思路：将最优剩余收益的减少率用贪心选择的边际收益下界，递推得到 $1-1/e$）。■

推论 1（顺序选择的自适应次模）
若选择下一阅读片段的决策依赖已观测事实（自适应环境），在“自适应次模”条件下，**自适应贪心**同样达成 $1-1/e$ 近似。

五、约束与拉格朗日对偶的等价

定义 9（软/硬约束）
引入约束向量 $\mathbf{g}(\mathcal{S}_T)\le \mathbf{b}$（如长度、引用覆盖率、一致性阈值）。考虑**约束 MDP**：

$$\max_\pi\ \mathbb{E}\big[\mathcal{V}(\mathcal{S}_T)\big]\quad \text{s.t. } \mathbb{E}\big[g_i(\mathcal{S}_T)\big]\le b_i,\ \forall i.$$

定理 4（拉格朗日等价与鞍点最优）
在有限状态/动作与 Slater 条件（存在严格可行解）下，原问题与对偶问题

$$\min_{\lambda\ge 0}\ \max_\pi\ \mathbb{E}\left[\mathcal{V}(\mathcal{S}_T)-\sum_i \lambda_i\big(g_i(\mathcal{S}_T)-b_i\big)\right]$$

零对偶间隙，存在鞍点 $(\pi^\star,\lambda^\star)$。

证明要点：约束 MDP 的线性规划表述+强对偶性（有限维情形）；π 上的极大与 λ 上的极小交换成立，得鞍点存在性。■

六、风险度量：CVaR 的等价表征与可优化性

定义 10（CVaR）
对随机变量 X（此处 $X=\mathcal{V}(\mathcal{S}_T)$ 的随机性来自策略与环境），置信水平 $\beta\in[0,1)$：

$$\mathrm{CVaR}_\beta(X)=\inf_{\eta\in\mathbb{R}}\ \eta+\frac{1}{1-\beta}\,\mathbb{E}\big[(X-\eta)_-\big].$$

定理 5（CVaR 的凸等价与可微替代目标）
CVaR_β 是一致（coherent）风险度量，上式给出其**凸优化等价**，从而在策略梯度/Actor-Critic 框架中可通过引入辅助变量 η 进行联合优化。

证明要点：Rockafellar–Uryasev 形式。凸合成与下期望保持凸性，故可联合最优化。■

七、POMDP → 信念 MDP 的严格化

定义 11（信念状态）
若认知资本存在隐藏成分或观测噪声，引入 $b_t \in \Delta(\mathcal{H})$（对潜在“应当被覆盖的要素”的分布），用 Bayes 更新：

$$b_{t+1} = \mathcal{T}(b_t, a_t, \mathrm{obs}_{t+1}).$$

定理 6（信念过程的马尔可夫性）
$\{b_t\}$ 在合适的 σ-代数上构成可测马尔可夫过程，并诱导**信念 MDP**，最优策略可在 $\pi(a \mid b)$ 中寻求。

证明要点：经典 POMDP→belief-MDP 构造；Bayes 预测-校正使未来仅依赖当前信念与动作。■

八、最大熵 RL 的软贝尔曼方程与 SAC 正当性

定义 12（软值与软 Q）
定义

$$V^\pi(s) = \mathbb{E}_{a\sim\pi}[Q^\pi(s,a) - \alpha \log \pi(a \mid s)], \quad Q^\pi(s,a) = \mathbb{E}\left[r(s,a) + \gamma V^\pi(s')\right].$$

定理 7（软贝尔曼算子的压缩与唯一性）
在有界奖励与 $\gamma \in (0,1)$ 下，软贝尔曼算子是 γ-压缩映射，存在唯一不动点 $(V^\star, Q^\star)$。

证明要点：用 sup-范数与 Jensen/Young 不等式，直接比照经典贝尔曼压缩；最大熵项保持算子单调并不破坏压缩。■

定理 8（Boltzmann 形最优策略）
存在最优策略 $\pi^\star$ 使

$$\pi^\star(a \mid s) \;\propto\; \exp\!\left(\tfrac{1}{\alpha} Q^\star(s,a)\right).$$

SAC 的策略更新（最小化 $\mathrm{KL}(\pi(\cdot \mid s) \,\|\, \frac{e^{Q^\star/\alpha}}{Z(s)})$）与评论家更新（最小化软贝尔曼残差）同时收敛到上述不动点附近。

证明要点：对固定 Q，极小化 $\mathbb{E}_{a\sim\pi}[-Q(s,a)+\alpha\log\pi]$ 得 Gibbs 分布；评论家用 TD/残差逼近不动点。■

九、蒸馏的性能差界（教师-学生）

定义 13（蒸馏与分布）
教师策略 π_T，学生 π_S。令训练分布 $d_{\pi_T}(s)$ 为教师诱导的状态访问分布。蒸馏目标：

$$\min_{\pi_S}\ \mathbb{E}_{s\sim d_{\pi_T}}\big[\mathrm{KL}\big(\pi_T(\cdot\mid s)\,\|\,\pi_S(\cdot\mid s)\big)\big]\le\varepsilon.$$

定理 9（性能差界，有限时域）
若对所有 s，$\|\pi_T(\cdot\mid s)-\pi_S(\cdot\mid s)\|_{\mathrm{TV}}\le\delta$（由 Pinsker：$\delta\le\sqrt{\tfrac{1}{2}\mathrm{KL}}$），且一步奖励范围直径为 $R_{\max}$，则长为 T 的期望回报差满足

$$|J(\pi_T)-J(\pi_S)|\ \le\ O(TR_{\max}\delta).$$

当 ε 较小，$\delta\le\sqrt{\varepsilon/2}$，故性能差 $=O(TR_{\max}\sqrt{\varepsilon})$。

证明要点：性能差分引理 + 访问分布偏差的 telescoping 分解 + Pinsker 不等式。■

十、总体正确性定理（“认知资管”管线）

定理 10（端到端正当性与最优性保证）
在上述设定下，以下结论成立：
(a)（**良定性**）最大熵目标的最优策略存在（定理 1）；
(b)（**策略不变性**）采用潜能塑形的稠密奖励不改变最优策略（定理 2）；
(c)（**近似最优性**）若估值函数为单调次模且受预算/成本约束，贪心/连续贪心达到 $1-1/e$ 近似（定理 3）；
(d)（**约束可解**）软/硬约束可经拉格朗日对偶等价求解并存在鞍点（定理 4）；
(e)（**风险稳健**）CVaR 目标可用凸等价进行可微优化（定理 5）；
(f)（**部分可观→可解**）POMDP 可经信念化转为 MDP 处理（定理 6）；
(g)（**算法正当**）SAC 的软值/软 Q 具有唯一不动点，策略为 Boltzmann 形（定理 7–8）；
(h)（**工程落地**）知识蒸馏的 KL 上界诱导有限时域性能差界（定理 9）。
综上，本框架把“长文阅读理解”的过程**严格地**嵌入一套可计算、可优化、可给出近似与稳定性保证的传统数学结构中。

十一、补充：从工程细节到数学假设的对齐

1. **代数结构选择**：若希望更强的合并/删冗不变性，可把 $(\mathbb{S}, \oplus)$ 取为**上半格**（join-semilattice），$\oplus$ 为并。事实图可赋以邻接集并。
2. **成本与滑点**：把上下文挤兑损失建模为次模函数的**负项**或显式成本项（保持总体可加与凸性）。
3. **可测性**：在可分度量空间（如嵌入向量空间的闭有界集）上，$\mathcal{F}$ 与 g 取 Borel 可测；策略取随机核即可。
4. **连续动作**：当动作含连续强度（如抽取/验证力度），软 Q 的压缩与存在性仍成立（在有界奖励与 $\gamma < 1$ 下）。

十二、结语（形式化的落地承诺）

这套严格化把你的直觉三件事**坐实**：

- 读是**做多覆盖、做空冗余、对冲不确定**的序贯交易；
- 估值可被**单调次模**封装，从而获得**近似最优保证**；
- 训练—蒸馏链路由**最大熵 RL + 对偶约束 + CVaR**支撑，具备**稳定性与稳健性**的传统数学后盾。

若你接着要“证明到代码”，可以把定理 2（塑形）、定理 3（次模）与定理 8（Boltzmann 策略）对应为三段最小可验的实验：潜能项不改排行、贪心基线给到 $1 - 1/e$ 近似、策略分布逐步靠近 Gibbs 形。这样，理论与工程将是可对拍的闭环。
