arXiv:2304.11641v1 [cs.FL] 23 Apr 2023

带有关于时序逻辑目标的优先级偏好的概率规划

Lening Li[1] , **Hazhar Rahmani**[2] , **Jie Fu**[2]

[1]Lening Li 就职于伍斯特理工学院机器人学系，地址为美国马萨诸塞州伍斯特市 01609，美国。[2]Hazhar Rahmani 和 Jie Fu 就职于佛罗里达大学电气与计算机工程系，地址为美国佛罗里达州盖恩斯维尔市 32605。lli4@wpi.edu，{h.rahmani、fujie}@ufl.edu *

摘要

本文研究概率环境中的时序规划问题，该环境被建模为带标签的马尔可夫决策过程（MDP），其中用户对多个时序目标具有偏好。现有研究将此类偏好表示为目标优先级列表。本文提出一种新的规范语言，称为有限轨迹上的优先定性选择线性时序逻辑。该语言在有限轨迹上的线性时序逻辑基础上，引入了优先定性选择逻辑中的优先合取和有序析取。借助该语言，可以简洁地规范时序目标，并表达完成每项时序任务时相应的偏好。描述系统行为的有限轨迹根据其相对于公式的不满意度分数进行排序。我们提出一种将该新语言系统地转换为加权确定性有限自动机的方法。利用这一计算模型，我们构建并求解一个问题：在给定用户偏好的情况下，计算使预期不满意度分数最小的最优策略。我们通过多个案例研究展示该逻辑和算法的有效性与适用性，并对每个案例进行了详细分析。

1 引言

在本研究中，我们研究给定时序目标偏好顺序的基于偏好的规划，即在时序逻辑中规定的有序目标。时序逻辑是用于规定复杂任务和任务目标的富有表现力且严谨的语言。带有时序逻辑目标的规划[Pnueli， 1981] 已被用于机器人系统 [Kantaros 等，2022年；Bradley 等，2021年；He 等，2020年；Vasile 等，2020年；Yang 等，2020年；Wang 等，2020年；Hekmatnejad 和 Fainekos，2018年；He 等，2015年；Li 等，2021] 以及其他智能系统 [Kasenberg 等，2020年；De Giacomo and Vardi，2015年；Camacho 等，2017年；Mallett 等，2021年；Zhou 等，2022年；Zhao 等， 2022]。

*本材料基于空军科学研究办公室资助的项目（授予编号 FA9550-21-1-0085），并部分获得美国国家科学基金会资助的项目（授予编号 2024802）。

规定时序目标上的偏好，使决策者能够灵活地修改任务，并在无法满足所有约束/子任务时实现最偏好的结果。早期研究考虑确定性系统——将其建模为有限离散系统，或具有确定性动力学的系统。一些研究 [Tumova 等，2013b；Tumova 等，2013a；Wongpiromsarn等，2021年；Vasile等， 2017] 提出了最小违反规划方法，用于决定在确定性系统中应违反哪些低优先级约束。 [Mehdipour 等，2021] 在信号时序逻辑中为布尔算子和时序算子关联权重，以规定满足子公式的重要性以及满足时序上的优先级。他们开发了算法，用于最大化确定性动力学系统中的加权满足度。 [Rahmani and O'Kane，2019年；Rahmani andO'Kane， 2020] 研究了同时给定目标硬规范和软规范时的时序规划，使用线性时序逻辑（LTL）和有限轨迹上的线性动态逻辑（LDL$_f$）。 [Cai 等， 2020] 考虑在最大化总奖励的同时，最小化与不可行的线性时序逻辑规范之间的偏差。

近年来，研究人员开始研究带有时序逻辑约束的基于偏好的概率规划。 [Li 等人， 2020]研究了受一系列按顺序排列的概率时序逻辑公式约束的马尔可夫决策过程（MDP）的基于偏好的规划。该算法按照优先级逐一枚举任务，直到找到满足最优先任务的策略。 [Lahijanian 和 Kwiatkowska2016] 研究了在由 MDP 建模的环境中处理不可行规范的语法协安全 LTL 规划。他们计算一个策略，使其最大化满足修订后的公式的概率，并最小化修订成本。 [Lac-erda 等人， 2015] 考虑了一个类似问题，其目标是按照优先级递减的顺序，最大化完成任务的概率，最大化向完成任务推进的概率，并最小化期望成本。

尽管已有研究探讨了基于概率偏好的规划，但人工智能中的偏好指定与时序目标上的偏好之间的联系尚未建立。我们提出了一种新语言，将模糊逻辑对偏好的表示融入时态逻辑。具体而言，我们考虑了由{v1}Brewka等人提出的定性选择逻辑（QCL）pro-

posed in [Brewka等人， 2004] 及其扩展——优先定性选择逻辑（PQCL） [Benferhat和Sedki，2007年; Benferhat和Sedki， 2008]。QCL通过一种称为有序析取的新逻辑连接词扩展了命题逻辑，记作 $\times$。公式 $A \times B$ 表示：如果可能，则选择 A；但如果A 不可能，则至少选择 B。PQCL通过引入优先级合取／析取扩展了QCL，使用户能够表达偏好中的优先级。结合起来，若 $(A \overrightarrow{\times} B)$ & $(C \overrightarrow{\times} D)$，其中&表示优先级合取，那么满足偏好 $A \times B$比满足偏好 $C \times D$更为重要$\overrightarrow{}$。

本文提出的偏好语言称为有限轨迹上的优先定性选择线性时序逻辑（PQCLTLf），它将 PQCL 与有限轨迹上的某个 LTL 子类相结合。具体而言，我们为原子偏好引入 LTL 公式，并利用 PQCL 表示对时序目标的偏好。根据 PQCL 的语义，我们为每个结果（状态时序）分配一个不满意度分数。该评分函数使我们能够构造一个基于偏好的概率规划目标，即在建模为带标签的马尔可夫决策过程的随机系统中，最小化期望不满意度分数。

然而，仅凭这种紧凑的逻辑偏好表示还不足以进行概率规划，因为概率规划通常需要一个计算模型。基于 LTL 与自动机之间的关系，我们开发了一种过程，可将优先定性选择时序逻辑（PQCTL）公式转换为加权确定性有限自动机。对于每条以程度 k 满足偏好的路径，该加权自动机都能确保其对应的加权自动机上的诱导运行的权重之和恰好为 k。通过使用乘积运算，将规划状态空间与加权自动机的状态集合相结合，我们证明，可以通过求解一个乘积马尔可夫决策过程，获得随机系统中最优的策略。该解的正确性取决于基于加权自动机中转移上的权重所定义的奖励函数。我们形式化地证明，在给定 PQCTL 公式的情况下，乘积马尔可夫决策过程中的奖励最大化策略能够使原始马尔可夫决策过程中的期望不满意程度最小化。在实验中，我们采用多个机器人运动规划示例来展示该方法的有效性和适用性，并详细比较基于偏好的规划与使用单体时序逻辑公式的传统概率规划。

2 预备知识

记号：有限集合 X 上所有概率分布的集合记为 $\mathcal{D}(X)$。

接下来介绍必要的预备知识和符号。

系统模型 我们将规划智能体（机器人）与其随机环境之间的交互建模为一种马尔可夫决策过程的变体。

定义 1 （带终止状态的标记马尔可夫决策过程）。带终止状态的标记马尔可夫决策过程（TLMDP）是一个元组 $M = \langle S, A := \cup_{s\in S} A_s, P, s_0, s_\perp, \mathcal{AP}, L\rangle$ 其中 S是有限状态集； A是有限动作集，对于每个状态 $s \in S$， A_s 是状态 s下的可用动作集； A 包含一个特殊的终止动作 $a_\perp$，且对于任意 $s \in S$， $a_\perp \in A_s$。 P: $S\times A\times S \to [0,1]$是转移概率函数，对于每个 s、 $s' \in S$和 $a \in A$， $P(s' \mid s, a)$ 表示马尔可夫决策过程在状态 s执行动作 a 后转移到 s' 的概率； $s_0 \in S$ 是初始状态； $s_\perp \in S$ 是终止状态，也是唯一的汇状态。对于任意 $s \in S$， $P(s, a_\perp, s_\perp) = 1$。也就是说，如果智能体选择终止动作 $a_\perp$，则必定能够到达终止状态 $s_\perp$。集合 $\mathcal{AP}$ 是有限的原子命题集； L: $S \to 2^{\mathcal{AP}} \cup \{\ltimes\}$ 是标记函数，它为每个状态 $s \in S \setminus \{s_\perp\}$赋予在 s成立的原子命题集合 $L(s) \subseteq \mathcal{AP}$。只有终止状态被标记为空字符串，即 $L(s_\perp) = \ltimes$。

此马尔可夫决策过程中的有限运行是一系列 $\varrho = s_0 a_0 s_1 a_1 \cdots s_{k-1} a_{k-1} s_k$，其中， s_0是初始状态，并且对于每个 $0 \le i \le k-1$， $P(s_{i+1} \mid s_i, a_i) > 0$。与此运行相关联的路径是序列 $\rho = s_0 s_1 \cdots s_k \in S^*$，该路径的轨迹定义为 $\text{trace}(\rho) = L(s_0)L(s_1)L(s_2)\cdots L(s_k) \in (2^{\mathcal{AP}})^*$。以 $s_\perp$ 结束的路径称为终止路径。

马尔可夫决策过程中的有限记忆随机策略是一个函数 π: $S^* \to \mathcal{D}(A)$，它将状态序列映射为动作分布。马尔可夫型或无记忆随机策略是马尔可夫决策过程中的一个函数 π: $S \to \mathcal{D}(A)$，它将当前状态映射为动作分布。我们将所有马尔可夫型随机策略的集合记为 Π。

有限记忆随机策略 π： $S^* \to \mathcal{D}(A)$按如下方式诱导出关于 S^*的马尔可夫链 $M^\pi = \langle S^*, P^\pi\rangle$： 对于任意 $\rho \in S^*, s \in S$,

$$P^\pi(\rho s|\rho) = \sum_{a\in A} P(s|\mathsf{Last}(\rho), a)\cdot \pi(\rho, a), \qquad (1)$$

其中，Last(ρ) 是给定序列 ρ的最后一个状态。

由马尔可夫策略诱导的随机过程是一个马尔可夫链 $M^\pi = \langle S^*, P^\pi\rangle$，其中 P^π可以作为（1）的特例获得。马尔可夫链 M^π中路径 ρ 的概率记为$\mathbf{Pr}(\rho;\ M^\pi)$。

规划目标 我们研究在一组时序目标上具有偏好的概率规划，这些目标用有限轨迹上的线性时序逻辑（LTLf）公式表示。

定义 2 （LTLf 语法 [De Giacomo and Vardi， 2013]）。给定一个有限集合 $\mathcal{AP}$ 原子命题，LTLf 公式的语法定义如下:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \bigcirc\varphi \mid \varphi\,\mathsf{U}\,\varphi,$$

其中 $p \in \mathcal{AP}$、否定（$\neg$）和合取（$\wedge$）是标准布尔算子，而“下一时刻”（$\bigcirc$）和“直到”（ U ）是时序算子。

非正式地说，公式 $\bigcirc\varphi$ 表示 φ 在下一时刻成立，而 $\varphi_1\,\mathsf{U}\,\varphi_2$ 表示存在一个未来时间点，使得 φ_2 成立，并且从当前时间到该未来时间点的所有时间瞬间， φ_1 都为真。时序算子“最终”（$\Diamond$）使用“直到”定义为 $\Diamond\varphi :=$ true $\mathsf{U}\,\varphi$。该算子的对偶算子是“总是”（$\Box$），其定义为

：$\Box\varphi := \neg\Diamond\neg\varphi$。公式 $\Diamond\varphi$ 表示存在某个未来时间点，使得 φ 成立，而 $\Box\varphi$ 则解释为 φ在所有未来时间点都为真。线性时序逻辑的语义f 被定义为有限轨迹上的解释，详见[De Giacomo and Vardi, 2013]。

LTLf 公式 φ的语言记作 $\mathcal{L}(\varphi)$，它是在字母表 $2^{\mathcal{AP}}$ 上满足 φ的有限轨迹集合。为简化记号，在下文中令 $\Sigma := 2^{\mathcal{AP}}$。有限字母表 Σ 上的所有有限词集合记作 Σ^*。LTLf 公式 φ 的语言可以表示为由确定性有限自动机（DFA）$\mathcal{A}_\varphi = \langle Q, \Sigma, \delta, q_0, F\rangle$接受的词集合，其中 Q 是有限状态集；$\Sigma = 2^{\mathcal{AP}}$ 是称为字母表的有限符号集；δ：$Q \times \Sigma \to Q$ 是转移函数，使得 $\delta(q, \sigma) = q'$表示从状态 q读取输入 σ 后到达的状态；$q_0 \in Q$是初始状态；$F \subseteq Q$ 是接受状态/终态集合。转移函数以通常方式递归扩展：对于给定的 $\sigma \in \Sigma$ 和 $w \in \Sigma^*$，$\delta(q, \sigma w) = \delta(\delta(q, \sigma)\ w)$。当且仅当 $\delta(q, w) \in F$时，词 w被接受。在给定LTLf语义的情况下，DFA$\mathcal{A}_\varphi$ 接受恰好满足 φ 的词集合。

3 偏好语言：优先级定性选择逻辑与时序逻辑的整合

在本节中，我们提出一种新的任务规范语言，用于描述时序目标上的一部分偏好。我们将这种语言称为 *PQCLTLf*，它将线性时序逻辑f 与 PQCL [Benferhat和Sedki, 2007]——一种用于表示排序目标的命题逻辑相结合。

$\overrightarrow{\text{PQCL}}$通过引入一个连接词 $\overrightarrow{\times}$（称为有序析取）扩展了命题逻辑：公式 $\phi_1 \overrightarrow{\times} \phi_2$ 表示，如果可能，则 ϕ_1；如果 ϕ_1 不可能，则 ϕ_2。该运算子 $\overrightarrow{\times}$ 是左结合的，因此 $\phi_1 \overrightarrow{\times} \phi_2 \overrightarrow{\times} \ldots \overrightarrow{\times} \phi_n = \phi_1 \overrightarrow{\times} (\phi_2 \overrightarrow{\times} (\ldots \overrightarrow{\times} \phi_n) \ldots\ldots)$。除有序析取外，PQCL 还引入了优先级合取：公式 $\phi_1 \& \phi_2$定义了 ϕ_1 和 ϕ_2的个体满足度之间的字典序。

定义 3（有限轨迹上的优先定性选择线性时序逻辑）。设 Φ 是定义在一组原子命题 $\mathcal{AP}$ 上的 LTLf 公式。定义在 $\mathcal{AP}$ 上的 PQCLTLf 片段（不含否定）为

$$\varphi := \psi \mid \varphi \overrightarrow{\times} \varphi \mid \varphi \& \varphi,$$

其中 $\psi \in \Phi$。

与 PQCL 相比，我们不包含否定运算，因此排除了优先析取。否定仅允许用于构造 LTLf 公式。之所以不包含否定，主要是因为存在歧义：对航班 $\times$火车的否定可能意味着这两个选项无差异、不可比较，或更偏好火车而非航班。

不同词语满足一个PQCLTLf 公式的程度各不相同，这促使我们量化一个PQCLTLf 公式可被满足的替代方案数量。

定义 4（可选性，扩展自 [Benferhat和Sedki, 2007]）。给定一个 PQCLTLf 公式 φ，φ 的可选性，记作 $opt(\psi)$，是 φ 可被满足的方式数量，其递归计算如下：

- If φ is an LTL$_f$ formula, then $opt(\varphi) = 1$;
- If $\varphi = \varphi_1 \overrightarrow{\times} \varphi_2$, then $opt(\varphi) = opt(\varphi_1) + opt(\varphi_2)$;
- If $\varphi = \varphi_1 \& \varphi_2$, then $opt(\varphi) = opt(\varphi_1) \cdot opt(\varphi_2)$.

与这种可选性的定义相关，对于每个词 $w \in \Sigma^*$ 和一个 PQCLTLf 公式，该词在某种满足度上满足该公式。

定义 5（满足度，扩展自 [Benferhat和Sedki, 2007]）。设 φ是定义在 $\mathcal{AP}$上的 PQCLTLf 公式，并且 $w \in \Sigma^*$（回顾 $\Sigma = 2^{\mathcal{AP}}$）是一个有限字。我们写作$w \models_k \varphi$，其中正整数 $k > 0$ 表示 w相对于 φ的满足度为 k；并使用 $w \not\models \varphi$表示 w 不满足 φ。

w 相对于 φ 的满足度计算如下：

- If φ is an LTL$_f$ formula, then $w \models_1 \varphi$ if $w \in \mathcal{L}(\varphi)$, and $w \not\models \varphi$ if $w \notin \mathcal{L}(\varphi)$.
- If $\varphi = \varphi_1 \overrightarrow{\times} \varphi_2$, then $w \models_k \varphi_1 \overrightarrow{\times} \varphi_2$ if either
  - $w \models_k \varphi_1$; or
  - $w \models_n \varphi_2$, $w \not\models \varphi_1$, and $k = n + opt(\varphi_1)$.
- If $\varphi = \varphi_1 \& \varphi_2$, then $w \models_k \varphi_1 \& \varphi_2$ if
  - there exist $i, j > 0$ such that $w \models_i \varphi_1$, $w \models_j \varphi_2$, and $k = opt(\varphi_2) \times (i-1) + j$;
  - otherwise, if $w \not\models \varphi_1$ or $w \not\models \varphi_2$, then $w \not\models \varphi_1 \& \varphi_2$.

满足度的定义仅对满足 PQCLTLf公式的所有词语集合建立了一个全序，但并未对不满足该公式的词语进行排序。出于规划目的，我们引入一个取值范围限定在 0 和 1 之间的度量。

定义 6（不满意度分数）。不满意度分数函数是一个函数 d：$\Sigma^* \times \Phi \to (0, 1]$，它为每个词语 $w \in \Sigma^*$ 和 PQCLTLf 公式 $\varphi \in \Phi$赋予区间 $(0, 1]$, 中的一个正实数，称为相对于 φ的不满意度分数，其计算如下：

- If $w \not\models \varphi$, then $\mathrm{d}(w, \varphi) = 1$;
- If $w \models_k \varphi$ for $k > 0$, then $\mathrm{d}(w, \varphi) = \frac{k}{opt(\varphi)+1}$.

分数越低，词语的满足度越高。请注意，该分数总是大于 0。在下文中，当公式 φ 可由上下文明确得知时，我们将不满意度分数 w 相对于 φ 简写为 $\mathrm{d}(w)$。

由 PQCLTL 公式生成的偏好模型f 公式每个定义在一组原子命题 $\mathcal{AP}$ 上的 PQCLTLf 公式 φ 都会在 $\Sigma^* = (2^{\mathcal{A}})^*$上诱导出一个偏好模型 $\succeq^\varphi$，使得对于任意两个词语 w，$w' \in \Sigma^*$，w 相对于 φ而言优于 w'，即 $w \succeq^\varphi w'$，当且仅当 $\mathrm{d}(w, \varphi) \leq \mathrm{d}(w', \varphi)$。

很容易证明以下性质。

引理 1。 如果 $w \succeq^{\varphi} w'$，则以下条件之一成立：1. $w \models_k \varphi$ 且 $w' \not\models \varphi$；或 2. $w \models_n \varphi$、 $w' \models_m \varphi$ 和$n \leq m$。

关于 Σ^* 的偏好模型可直接转换为关于 S^*的偏好模型——即带标签的马尔可夫决策过程中有限路径的集合，其中路径 $\rho \in S^*$ 偏好于 $\rho' \in S^*$，当且仅当 $\mathsf{d}(L(\rho)\ \varphi) \leq \mathsf{d}(L(\rho')\ \varphi)$。因此，给定带标签的马尔可夫决策过程中的一个策略，我们引入以下度量，以评估相对于 PQCLTL$_f$ 公式而言该策略的偏好程度。

定义 7（期望不满意度分数）。 设 π 是给定 MDP 的有限记忆随机策略， $M^{\pi} = \langle S^*, P^{\pi} \rangle$ 是其诱导马尔可夫链， φ 是一个 PQCLTL$_f$公式。 π 相对于 φ的期望不满意度分数记为 $\mathsf{d}(\pi,\ \varphi)$，定义为

$$\mathsf{d}(\pi, \varphi) = \sum_{\rho \in S^*} \mathbf{Pr}(\rho; M^{\pi}) \cdot \mathsf{d}(L(\rho), \varphi). \tag{2}$$

现在，我们正式陈述概率规划问题：

问题：具有优先级偏好的时序逻辑目标概率规划（PPwPPoTLO）
输入：一个带标签的马尔可夫决策过程 $M = \langle S, A := \cup_{s \in S} A_s,\ P,\ s_0,\ s_{\perp},\ \mathcal{AP},\ L \rangle$ 以及一个 PQCLTL$_f$公式 φ。输出：一个策略 π: $S^* \to \mathcal{D}(A)$，使 φ的期望不满意度分数最小。

4 PQCLTL$_f$ 公式的最优规划

现在，我们提出一种规划算法来解决 PPwPPoTLO 问题。我们的方法分为两个步骤：第一步，我们为 PQCLTL$_f$ 公式构建一个自动机理论模型。第二步，我们证明，可以通过求解一个具有增广状态的奖励最大化马尔可夫决策过程，计算出使给定公式的期望不满意度分数最小的最优策略。

4.1 PQCLTL$_f$公式的自动机理论建模

我们重点关注如何为给定的 PQCTL 公式 φ 构建计算模型。用于表示 PQCLTL$_f$ 公式这一子类的计算模型选择为加权确定性有限自动机。

定义 8（加权确定性有限自动机 [Droste 和 Gastin， 2009]）。加权确定性有限自动机是一个元组 $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, \mathbf{w} \rangle$，其中 Q 是有限状态集； $\Sigma \cup \{\ltimes\}$ 是有限符号集（字母表）； $\ltimes$ 是表示字符串结尾的唯一符号[1]； δ: $Q \times (\Sigma \cup \{\ltimes\}) \to Q$ 是确定性转移函数； q_0 是初始状态； $\mathbf{w}$: $Q \times (\Sigma \cup \{\ltimes\}) \times Q \to \mathbb{R}$是权重函数，它将每个转移 (q, σ, q') 映射为一个实数，称为该转移的权重。

[1]通常，可以将 $\rtimes$ 作为有限字符串的起始部分，将 $\ltimes$ 作为有限字符串的结尾部分。起始符号 $\rtimes$ 因可根据上下文明确判断而省略。

考虑一个有限字 $w = \sigma_0\sigma_1 \ldots \sigma_{n-1}\ltimes$，令 $w[i]$ 为该字的第 i 个符号。由字w 生成的运行 ρ 为 ρ: $= q_0\sigma_0 q_1 \ldots \sigma_{n-1} q_n$，满足 $q_{i+1} = \delta(q_i,\ w[i])$，其中 $i = 0, \ldots,\ n-1$。我们用 $\mathrm{Word}(\rho) = w$ 表示与运行 ρ 相关联的字。总权重为 $\mathbf{w}(\rho) = \sum_{i=0}^{n-1} \mathbf{w}(q_i,\ w[i],\ q_{i+1})$。

首先，我们展示如何为 LTL$_f$ φ 构建加权确定性有限自动机（WDFA）。

定义 9（用于 LTL$_f$ 公式 φ 的 WDFA）。 设 $\mathcal{A}_{\varphi} = \langle Q, \Sigma, \delta, q_0, F \rangle$ 是对 φ 进行编码的确定性有限自动机。用于编码φ 的 WDFA 可由 $\mathcal{A}_{\varphi}$ 构造为一个元组

$$\mathcal{A} = \langle Q \cup \{\mathsf{sink}\}, \Sigma \cup \{\ltimes\}, \delta', q_0, \mathbf{w} \rangle$$

其中，对于每个 $q \in Q \cup \{\text{汇点}\}$ 和 $\sigma \in \Sigma \cup \{\ltimes\}$，

$$\delta'(q, \sigma) = \begin{cases} \delta(q, \sigma) & \text{if } q \neq \mathsf{sink} \text{ and } \sigma \neq \ltimes \\ \mathsf{sink} & \text{otherwise,} \end{cases} \tag{3}$$

并且，对于每个 $q,\ q' \in Q \cup \{\text{汇点}\}$ 和 $\sigma \in \Sigma \cup \{\ltimes\}$，

$$\mathbf{w}(q, \sigma, q') = \begin{cases} 1 & \text{if } q \in F \text{ and } \sigma = \ltimes \text{ and } q' = \mathsf{sink} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

直观地说，WDFA $\mathcal{A}$ 通过添加一个汇状态 sink 扩展了 DFA $\mathcal{A}_{\varphi}$。对于原 DFA $\mathcal{A}_{\varphi}$ 的任意状态 $s \in S$，在输入符号 $\ltimes$ 下会转移到 sink。只有当读取结束符号 $\ltimes$ 时，转移从接受状态到汇状态，才会获得权重 1。

引理 2。 给定一个 LTL$_f$ 公式 φ 的 WDFA $\mathcal{A}$ 和一个有限运行 $\rho = q_0\sigma_0 q_1 \ldots \sigma_{n-1} q_n$，如果 $\mathbf{w}(\rho) = 1$，则 $\mathrm{Word}(\rho) \models_1 \varphi$。

引理 2、3 4和5 的证明可见附录 A。

接下来，我们定义 PQCLTL$_f$公式的有序析取和优先级合取的 WDFA 构造过程。

定义 10（PQCLTL$_f$公式的有序析取的 WDFA）。设 $\mathcal{A}_i = \langle Q_i \cup \{\mathsf{sink}_i\}、\ \Sigma \cup \{\ltimes\}、\ \delta_i、\ q_{0_i}、\ \mathbf{w}_i \rangle$分别为 $i = 1$、2 个 WDFA，它们分别编码两个 PQCLTL$_f$ 公式 φ_1、 φ_2。可以由它们构造出一个表示 $\varphi_1 \overset{\rightarrow}{\vee} \varphi_2$ 的 WDFA，其元组为 $\mathcal{A} = \langle Q_1 \times Q_2 \cup \{\mathsf{sink}\}、\ \Sigma \cup \{\ltimes\}、\ \delta、\ (q_{0_1},\ q_{0_2})、\ \mathbf{w} \rangle$，其中，转移函数定义如下：对于任意 $(q_1,\ q_2) \in Q_1 \times Q_2$ 和 $\sigma \in \Sigma \cup \{\ltimes\}$，

$$\delta((q_1, q_2), \sigma) = \begin{cases} (\delta_1(q_1, \sigma), \delta_2(q_2, \sigma)) & \text{if } \sigma \neq \ltimes \\ \mathsf{sink} & \text{otherwise,} \end{cases}$$

且权重函数定义为：

- For any $(q_1, q_2) \in Q_1 \times Q_2$, input $\sigma \in \Sigma$
$$\mathbf{w}((q_1, q_2), \sigma, (\delta_1(q_1, \sigma), \delta_2(q_2, \sigma))) = 0$$
- For any $(q_1, q_2) \in Q_1 \times Q_2$, input $\ltimes$,
$$\mathbf{w}((q_1, q_2), \ltimes, \mathsf{sink}) = \begin{cases} \mathbf{w}_1(q_1, \ltimes, \mathsf{sink}) & \text{if } \mathbf{w}_1(q_1, \ltimes, \mathsf{sink}) > 0 \\ \mathbf{w}_2(q_2, \ltimes, \mathsf{sink}) + opt(\varphi_1) & \text{if } \mathbf{w}_1(q_1, \ltimes, \mathsf{sink}) = 0 \\ & \text{and } \mathbf{w}_2(q_2, \ltimes, \mathsf{sink}) > 0 \\ 0 & \text{otherwise,} \end{cases}$$

→**引理 3**。给定一个关于 $\varphi = \varphi_1 \times \varphi_2$ 的 WDFA $\mathcal{A}$ 和一个有限运行 $\rho = \mathbf{q}_0\sigma_0\mathbf{q}_1 \dots \mathbf{q}_{n-1}\sigma_{n-1}\mathbf{q}_n$，如果对于一个 $k > 0$，$\mathbf{w}(\rho) = k$，则 Word(ρ) $\models_k \varphi$，否则 Word(ρ) $\not\models \varphi$。

定义 11（用于 PQCLTL$_f$ 公式的优先级合取的 WDFA）。设 $\mathcal{A}_i = \langle Q_i \cup \{\text{汇点}_i\}, \Sigma \cup \{\ltimes\}, \delta_i, q_{0i}, \mathbf{w}_i\rangle$ for $i = 1$ 和 2 是分别编码两个 PQCLTL$_f$ 公式 φ_1、φ_2 的两个 WDFA。可以由它们构造出一个用于 $\varphi_1 \& \varphi_2$ 的 WDFA，其元组表示为 $\mathcal{A} = \langle Q_1 \times Q_2 \cup \{\text{汇点}\}, \Sigma \cup \{\ltimes\}, \delta, (q_{01}, q_{02}), \mathbf{w}\rangle$，其中，转移函数定义为：对于任意 $(q_1, q_2) \in Q_1 \times Q_2$和 $\sigma \in \Sigma \cup \{\ltimes\}$,

$$\delta((q_1,q_2),\sigma) = \begin{cases} (\delta_1(q_1,\sigma), \delta_2(q_2,\sigma)) & \text{if } \sigma \neq \ltimes, \\ \mathsf{sink} & \text{otherwise.} \end{cases}$$

权重函数定义为:

- For any $(q_1, q_2) \in Q_1 \times Q_2$, for $\sigma \in \Sigma$,
$$\mathbf{w}((q_1,q_2),\sigma,(\delta_1(q_1,\sigma),\delta_2(q_2,\sigma))) = 0$$
- 对于任意 $(q_1, q_2) \in Q_1 \times Q_2$，对于输入 $\ltimes$，如果 $\mathbf{w}_i(q_i, \ltimes, \text{汇点}) > 0$ for both $i = 1, 2$，则
$$\mathbf{w}((q_1,q_2),\ltimes,\mathsf{sink}) = \mathbf{w}_2(q_2,\ltimes,\mathsf{sink}) + opt(\varphi_2)\cdot(\mathbf{w}_1(q_1,\ltimes,\mathsf{sink}) - 1),$$
else $\mathbf{w}((q_1,q_2),\ltimes,\mathsf{sink}) = 0$.

引理 4. 给定一个用于 $\varphi = \varphi_1 \& \varphi_2$ 的加权确定性有限自动机 $\mathcal{A}$，以及一个有限运行 $\rho = \mathbf{q}_0\sigma_0\mathbf{q}_1 \dots \mathbf{q}_{n-1}\sigma_{n-1}\mathbf{q}_n$，若 $\mathbf{w}(\rho) = k$ 对于某个 $k > 0$ 成立，则 Word(ρ) $\models_k \varphi$；否则，Word(ρ) $\not\models \varphi$。

根据上述针对 PQCLTL$_f$ 公式的加权确定性有限自动机构造方法，可以递归地构造针对更复杂的 PQCLTL$_f$ 公式的加权确定性有限自动机。

引理 5 $_f$ φ $\mathcal{L}(\varphi) \neq \emptyset$。给定一个满足 $\mathcal{A}$() 的 PQCLTL 公式及其构造出的加权确定性有限自动机，φ 的可选性是 $\mathcal{A}$中所有转移的最大权重。即

$$opt(\varphi) = \max\{\mathbf{w}(q,a,q') \mid \delta(q,a,q') \text{ is defined.}\}$$

附录B中给出了一个示例，用于说明加权确定性有限自动机的构造。

5 以最小化不满意度分数为目标的概率规划

本节将展示如何利用加权确定性有限自动机求解 PPwPPoTLO 问题。与带有线性时序逻辑约束的概率规划类似，在带标签的马尔可夫决策过程和加权确定性有限自动机之间执行乘积运算，可以帮助我们跟踪时序目标。

定义 12（带标签的马尔可夫决策过程与 WDFA 的乘积）。给定的 WDFA $\mathcal{A} = \langle Q \cup \{\text{汇点}\}, \Sigma \cup \{\ltimes\}, \delta, q_0, \mathbf{w}\rangle$ 与终止型带标签的马尔可夫决策过程 $M = \langle S, A := \cup_{s\in S} A_s, P, s_0, s_\perp, \mathcal{AP}, L\rangle$ 的乘积是一个马尔可夫决策过程

$$\mathcal{M} = M \otimes \mathcal{A} = (V, A := \bigcup_{v\in V} A_v, \mathcal{P}, v_0, R)$$

其中

- $V = S \times Q$ is the state space,
- A 是动作集合，对于每个 $v = (s,q) \in V$，$A_v = A_s$ 是 v 的可用动作集合,
- $\mathcal{P}$ 是概率转移函数，其中对于每个状态（s、q）、（s'、q'）$\in V$和动作 $a \in A$,
$$\mathcal{P}((s,q),a,(s',q')) = P(s,a,s')\cdot\mathbf{1}(\delta(q,L(s')) = q')$$
- $v_0 =$（s_0、δ（q_0、L（s_0））））是初始状态。
- R：$V \times A \to \mathbb{R}$ 是奖励函数，其中对于每个（s、q）$\in V$和 $a \in A$，如果 $a = a_\perp$且$\mathbf{w}$（q、$\ltimes$、汇点）> 0，则 R（（s、q）$a_\perp$）$= opt$（φ）$- \mathbf{w}$（q、$\ltimes$、汇点）$+ 1$，否则R（（s、q）a）$= 0$。

给定乘积马尔可夫决策过程中的一个有限运行 $h = v_0a_0v_1a_1\dots v_n$，总奖励为 $R(h) = \sum_{i=0}^{n-1} R(v_i, a_i)$。由于乘积马尔可夫决策过程中的运行 h 对应于原始马尔可夫决策过程中的运行 ρ，区别仅在于 ρ 中的每个状态都附加了一个自动机状态，因此我们使用 Proj$s(h)$ 来计算运行 $h = (s_0, q_0)(s_1, q_1)\dots(s_n, q_n) \in V^*$ 在运行 $s_0s_1s_2\dots s_n \in S^*$ 上的投影，其标记为 $L(s_0s_1\dots s_n) = L(s_0)L(s_1)\dots L(s_n)$。我们用 $Runs(\mathcal{M})$ 表示 $\mathcal{M}$中的有限运行集合。

根据奖励函数，初始状态$v \in V$下非平稳策略 $\pi : V^* \to \mathcal{D}(A)$ 的期望总奖励定义为

$$J_\pi(v) = \limsup_{N\to\infty} J_{\pi,N}(v),$$

其中，$J_{\pi,N}(v)$ 表示 π状态 v 的期望 N 阶段奖励:

$$J_{\pi,N}(v) = E\left[\sum_{t=0}^{N} R(V_k, \pi(V_0\dots V_k)) \mid V_0 = v\right],$$

其中，V_k 是时间 k 的状态。该期望是相对于马尔可夫链 $\mathcal{M}^\pi$ 中路径的概率分布而言的。

引理 6。对于乘积马尔可夫决策过程 $\mathcal{M}$ 的任意策略 $\pi : V^* \to \mathcal{D}(A)$，对于任意 $v \in V$、$J_\pi(v) < \infty$。

证明见附录 A。

最优价值函数定义为

$$J^*(v) = \arg\max_\pi J_\pi(v).$$

为了通过最优规划最大化总奖励，$J^*(v)$ 可以通过马尔可夫策略 [Puterman, 2014] 实现。因此，下面我们仅考虑马尔可夫策略。我们还考虑贝尔曼算子 T，其定义为

$$TJ(v) = \max_{\pi\in\Pi} \sum_{a\in A_v} [R(v,a)\pi(a \mid v) + \sum_{v'\in V} \mathcal{P}(v'|v,a)\pi(a \mid v)J(v')],$$

且最优价值函数满足 $TJ^* = J^*$。

在乘积马尔可夫决策过程的所有马尔可夫策略中，我们仅考虑适当 的策略。

定义13（扩展自 [Bertsekas 和 Yu, 2013]）。马尔可夫决策过程 π 的策略 $\mathcal{M}$ 是 适当的，如果它能保证以概率为1到达汇状态（$s_\perp$，汇点）。

引理 7。对于任意 $v \in V$，马尔可夫决策过程乘积 $\mathcal{M}$ 的最优值 $J^*(v)$ 可由一个适当的马尔可夫策略得到。

证明见附录A。

因此，为了搜索最优策略，我们只需考虑适当的马尔可夫策略。现在，我们将乘积马尔可夫决策过程中的奖励最大化问题与最小化期望不满意度分数的规划目标联系起来。

引理 8。对于每条路径 $h = (s_0, q_0)(s_1, q_1) \cdots (s_n, q_n) \in V^*$，均有：

$$\mathsf{d}(L(\mathsf{Proj}_S(h)), \varphi) = \begin{cases} 1 - \frac{R(h)}{opt(\varphi)+1} & \text{if } R(h) > 0, \\ 1 & \text{if } R(h) = 0. \end{cases}$$

证明见附录 A。

定理 1。设 $\pi : V \to \mathcal{D}(A)$ 是乘积马尔可夫决策过程 $\mathcal{M}$ 的一个策略。由 π 构造策略 $\pi' : S^* \to \mathcal{D}(A)$，使得对于每个 $\rho : s_0 s_1 \cdots s_n \in S^*$, $\pi'(\rho) = \pi((s_n, \delta(q_0, L(\rho))))$。如果 π 是 $\mathcal{M}$ 的最优策略，那么 π' 是使期望不满意度分数最小化的最优策略，即 PPwPPoTLO 的解。

证明。我们建立 π' 的期望不满意度分数与 π的值之间的联系。首先，我们使用（2）展开 π'的期望不满意度分数：

$$\mathsf{d}(\pi', \varphi) = \sum_{\rho \in S^*} \mathbf{Pr}(\rho; M^{\pi'}) \cdot \mathsf{d}(L(\rho), \varphi) \tag{5}$$

接下来，我们展开 π的值。

$$\begin{aligned} J_\pi(v_0) &= \sum_{h \in V^*} \mathbf{Pr}(h; \mathcal{M}^\pi) \cdot R(h) \\ &= \sum_{h \in V^* : R(h) = 0} \mathbf{Pr}(h; \mathcal{M}^\pi) \cdot 0 \\ &+ \sum_{h \in V^* : R(h) \neq 0} \mathbf{Pr}(h; \mathcal{M}^\pi) \cdot R(h) \end{aligned} \tag{6}$$

利用引理 8的结果，我们将该求和写为：

$$\begin{aligned} J_\pi(v_0) = & \sum_{h \in V^* : R(h) \neq 0} \mathbf{Pr}(h; \mathcal{M}^\pi) \cdot (1 - \mathsf{d}(L(\mathsf{Proj}_S(h)))) \\ & \cdot (opt(\varphi) + 1)) \\ & [\textit{ replace } opt(\varphi) + 1) \textit{ by } K, \mathbf{Pr}(h; \mathcal{M}^\pi) \textit{ by } \mathbf{Pr}^\pi(h).] \\ =& (K \cdot \sum_{h \in V^* : R(h) \neq 0} \mathbf{Pr}^\pi(h) \\ & - K \cdot \sum_{h \in V^* : R(h) \neq 0} \mathbf{Pr}^\pi(h) \mathsf{d}(L(\mathsf{Proj}_S(h))) \\ =& K \cdot \sum_{h \in V^*} \mathbf{Pr}^\pi(h) - K \cdot \sum_{h \in V^* : R(h) = 0} \mathbf{Pr}^\pi(h) \\ & - K \cdot \sum_{h \in V^* : R(h) \neq 0} \mathbf{Pr}^\pi(h) \mathsf{d}(L(\mathsf{Proj}_S(h))) \qquad (7) \\ =& (K - K \cdot \sum_{h \in V^* : R(h) = 0} \mathbf{Pr}^\pi(h) \cdot 1 \\ & - K \cdot \sum_{h \in V^* : R(h) \neq 0} \mathbf{Pr}^\pi(h) \mathsf{d}(L(\mathsf{Proj}_S(h))) \qquad (8) \\ =& K - K \cdot \sum_{h \in V^*} \mathbf{Pr}^\pi(h) \cdot \mathsf{d}(L(\mathsf{Proj}_S(h))) \qquad (9) \end{aligned}$$

从（7）到（8），我们使用概率公理：$\sum_{h \in V^*} \mathbf{Pr}(h); \mathcal{M}^\pi) = 1$。从（8）到（9），我们使用引理 8：如果 $R(h) = 0$ 则 $\mathsf{d}(L(\mathsf{Proj}_S(h))) = 1$。因此，结合（9）和（5），我们得到

$$J_\pi(v_0) = K - K \cdot \mathsf{d}(\pi', \varphi), \tag{10}$$

因此，$\mathrm{argmax}_\pi J_\pi(v_0) = \mathrm{argmin}_{\pi'} \mathsf{d}(\pi', \varphi)$，也就是说，使 J 最大化的策略 π 会得到使不满意度分数 d 最小化的策略 π'。□

6 复杂度分析

算法的第一步是构造一个对 φ进行编码的加权确定性有限自动机。从线性时序逻辑f 公式构造出的确定性有限自动机，在最坏情况下其规模相对于公式大小呈双指数增长 [Wolper, 2001; DeGiacomo 和 Favorito, 2021]。然而，在实践中，这种转换对于机器人规划中常见的线性时序逻辑f 公式是可处理的。有序析取自动机的构造 $\varphi_1 \to \varphi_2$ 以及优先级合取 $\varphi_1 \& \varphi_2$ 使用定义10 和定义11，其所需时间相对于子公式 φ_1 和 φ_2对应的 WDFA 的规模为多项式时间。构造乘积马尔可夫决策过程 $\mathcal{M}$ 所需时间相对于 WDFA 和原始 MDP 的规模为多项式时间。使用标准技术（值迭代/策略迭代或线性规划）计算 $\mathcal{M}$ 的最优策略所需时间相对于乘积 MDP 的规模为多项式时间 $\mathcal{M}$。

7 实验

我们通过若干概率机器人运动规划示例展示所提出算法的有效性。[2]

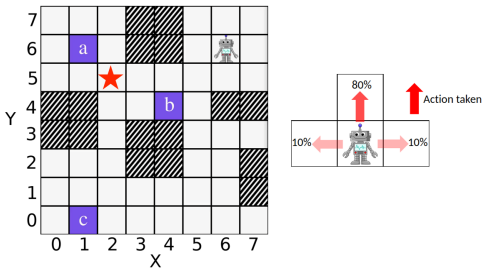


图 1：一个 8×8 随机网格世界 g_1（不含红色星标）以及执行动作“N”时的转移概率。红色星标是在网格世界 g_2中额外引入的一个洞。

考虑一个小型随机网格世界 g_1，如图1所示。对于每个状态 $s \in S$，机器人有四个动作：“N”、“W”，

[2]所有实验均在一台配备 AMD Ryzen 9 5900X CPU 和 32 GB 内存的 Ubuntu 20.04 机器上执行。我们使用 Gurobi 求解器对马尔可夫决策过程进行规划。对于具有不同公式的任意 8×8 网格世界，求解最优规划问题的计算时间均不超过 0.1 秒。代码可在补充文件中找到。

“S”、“E”。机器人从某个状态执行动作后，以 0.8 的概率转移到预期的网格单元，以 0.1 的概率滑入非预期的网格单元。如果机器人执行动作后到达边界墙，则停留在原网格单元中。机器人的初始状态为 (6,6)。阴影区域表示洞。一旦机器人进入洞中，就会被困住。感兴趣区域标记为 a、b 和 c。相应地，$\mathcal{AP} = \{a, b, c\}$。当机器人位于以相应原子命题标记的区域中时，这些原子命题中的每一个均在某个时间瞬间成立。

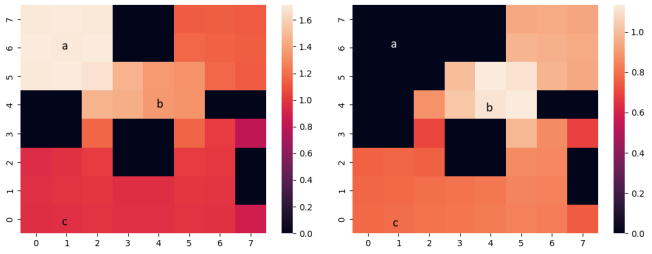
给定原子命题集合 $\mathcal{AP}$，我们考虑以下偏好公式（WDFA 见附录 B。）

$$\Diamond b \overrightarrow{\times} (\Diamond a \vee \Diamond c).$$

我们计算了使期望不满意度分数最小化的最优策略 π_*。为了观察有序析取与常规析取的差异，我们还计算了一个使满足公式 $\Diamond b \vee (\Diamond a \vee \Diamond c)$ 的概率最大化的最优策略。我们将此策略记为 $\pi_\vee$。图 2b 展示了不同初始状态下的最优值。随后，我们分别对 π_* 和 $\pi_\vee$ 针对 $\Diamond b$ 和 $\Diamond a \vee \Diamond c$ 进行策略评估。图 2c、2d、2e 和 2f 展示了不同初始状态下 $\phi \in \{\Diamond b, \Diamond a \vee \Diamond c\}$ 满足公式 ϕ 的概率。比较公式 $\Diamond b$ 对应的图 2c 与图 2e 可见，π_* 在网格世界的大部分区域取得了更高的值，尤其是在左上角。另一方面，比较图 2d 与图 2f 可见，$\pi_\vee$ 在网格世界的大部分区域满足 $\Diamond a \vee \Diamond c$ 的概率高于策略 π_*，尤其是在右上角。这一比较表明，当 $\Diamond b$ 优先于 $\Diamond a \vee \Diamond c$ 时，基于偏好的策略会倾向于满足 $\Diamond b$。

下一时刻，我们考虑以下具有优先级合取和嵌套有序析取的公式：$\varphi_3 = \varphi_1 \& \varphi_2$，其中 $\varphi_1 = \Diamond b \overrightarrow{\times} (\Diamond a \vee \Diamond c)$ 和 $\varphi_2 = \Diamond (a \wedge \Diamond (b \wedge \Diamond c)) \overrightarrow{\times} \Diamond (a \wedge \Diamond c) \vee \Diamond (b \wedge \Diamond c)$。该任务公式描述了系统需要同时满足 φ_1 和 φ_2，其中φ_1 的优先级高于 φ_2。

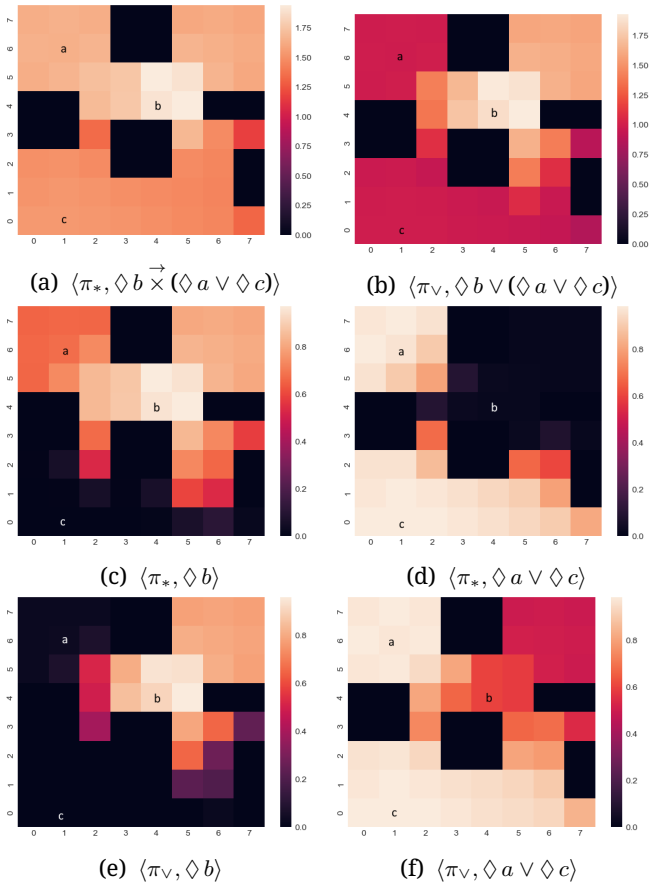
对于此情形，我们考虑一个额外的网格世界 g_2，其中在位置 (2，5) 增加了一个额外的洞，该洞阻断了通往 a 的通路。给定公式 φ_3，我们计算区域 a 可通行时的最优策略 $\pi_\ddagger$，以及区域 a 不可通行时的最优策略 $\pi_\ddagger^-$。



(a) (b)

图3：(a) 网格世界 g_1 中的值 $J((\cdot, \mathbf{q}_0); \langle \pi_\ddagger, \varphi_3 \rangle)$。(b) 网格世界 g_2 中的值 $J((\cdot, \mathbf{q}_0); \langle \pi_\ddagger^-, \varphi_3 \rangle)$。

我们绘制了这两个网格世界中 φ_3 的热力图，见



(a) $\langle \pi_*, \Diamond b \overrightarrow{\times} (\Diamond a \vee \Diamond c) \rangle$ (b) $\langle \pi_\vee, \Diamond b \vee (\Diamond a \vee \Diamond c) \rangle$

(c) $\langle \pi_*, \Diamond b \rangle$ (d) $\langle \pi_*, \Diamond a \vee \Diamond c \rangle$

(e) $\langle \pi_\vee, \Diamond b \rangle$ (f) $\langle \pi_\vee, \Diamond a \vee \Diamond c \rangle$

图 2：每个带有子图标题 $\langle \pi, \phi \rangle$ 的子图表示在网格世界 g_1 中，给定公式 ϕ、从不同初始状态开始时，对策略 π 进行策略评估所得的值 $J((\cdot, \mathbf{q}_0); \langle \pi, \phi \rangle)$。

图3a 和 3b。我们观察到：当a 可通行时，从左上角开始，智能体在最优策略下获得的值更高。但如果 a 不可通行，左上角的状态值为零。这是因为公式 φ_2 无法满足，因为智能体从左上角开始时无法到达区域 c。因此，φ_3 不可满足。在给定 a 可通行的情况下，最优策略下的状态值高于 a 不可通行时的状态值，这表明智能体可以在网格世界中实现更受偏好的结果 g_1。

8 结论

本文介绍了一种新的规范语言，称为有限轨迹上的优先级定性选择线性时序逻辑（PQCLTLf），用于紧凑地描述时序目标以及用户对子目标的偏好。我们给出了从该语言到加权确定性有限自动机的自动转换方法。我们利用这一转换解决了如下问题：在由马尔可夫决策过程（MDP）建模的随机环境中，计算一个策略，使给定 PQCTL 公式的不满意度分数期望最小。通过弥合人工智能中的偏好与时序逻辑规划之间的鸿沟，本工作为未来研究纳入偏好 elicitation、从正/负数据中学习以及序列中的自适应规划提供了可能

决策问题。

参考文献

[Benferhat和Sedki， 2007] Salem Benferhat和Karima Sedki。处理优先级偏好的修订版定性选择逻辑。载于《符号与定量推理及不确定性方法欧洲会议》，第635–647页。Springer，2007年。[Benferhat和Sedki， 2008] Salem Benferhat和Karima Sedki。处理定性选择逻辑中偏好的两种替代方案。《模糊集与系统》，159(15)：1889–1912，2008年8月。[Bertsekas 和 Yu， 2013] Dimitri P Bertsekas和Huizhen Yu。弱条件下的随机最短路径问题。信息与决策系统实验室报告 LIDS-P-2909，麻省理工学院，2013。[Bradley等人， 2021] Christopher Bradley、Adam Pacheck、Gregory J Stein、Sebastian Castro、Hadas Kress-Gazit和Nicholas Roy。在未知环境中针对时间扩展任务的学习与规划。载于2021年 IEEE 机器人与自动化国际会议（ICRA），第4830–4836页。IEEE，2021年。[Brewka等人， 2004] Gerhard Brewka、Salem Benferhat和Daniel Le Berre。定性选择逻辑。《人工智能》，157(1)：203–237，2004年8月。[Cai等人， 2020] Mingyu Cai、Hao Peng、Zhijun Li、Hongbo Gao和Zhen Kan。基于滚动时域控制的运动规划与部分不可满足的LTL约束。IEEE 控制系统快报，5(4)：1279–1284，2020年。[Camacho等人， 2017] Alberto Camacho、Eleni Triantafil- lou、Christian Muise、Jorge A Baier和Sheila A McIl- raith。带有时间扩展目标的非确定性规划：有限轨迹和无限轨迹上的LTL。载于第三十一届 AAAI 人工智能会议，2017。[De Giacomo and Favorito， 2021] Giuseppe De Giacomo和Marco Favorito。将LTLf/LDLF转换为确定性有限自动机的组合方法。载于《国际自动规划与调度会议论文集》，第31卷，第122–130页，2021年。[De Giacomo and Vardi， 2013] Giuseppe De Giacomo和Moshe Y Vardi。有限轨迹上的线性时序逻辑和线性动态逻辑。载于 IJCAI'13 第二十三届国际人工智能联合会议论文集，第854–860页。计算机协会，2013。[De Giacomo and Vardi， 2015] Giuseppe De Giacomo和Moshe Vardi。有限轨迹上的LTL和LDL综合。载于第二十四届国际人工智能联合会议，2015。[Droste 和 Gastin， 2009] Manfred Droste和Paul Gastin。加权自动机与加权逻辑。载于《加权自动机手册》，第175–211页。Springer，2009年。[He等人， 2015] Keliang He、Morteza Lahijanian、Lydia E Kavraki和Moshe Y Vardi。面向带时序逻辑规范的操作规划。载于2015 *IEEE*机器人与自动化国际会议（ICRA），第 346–352 页。IEEE，2015年。[He 等， 2020] Binghan He、Jaemin Lee、Ufuk Topcu 和 Luis Sentis。BP-RRT：面向时序逻辑运动规划的屏障对综合。载于 2020年 IEEE 第 59 届决策与控制会议（CDC），第 1404–1409 页。IEEE，2020年。[Hekmatnejad 和 Fainekos， 2018] Mohammad Hekmatnejad 和 Georgios Fainekos。具有访问权限级别的系统的最优多值 LTL 规划。载于 2018年美国控制会议（ACC），第 2363–2370 页。IEEE，2018年。[Kantaros 等， 2022] Yiannis Kantaros、Samarth Kalluraya、Qi Jin 和 George J Pappas。不确定语义地图中的基于感知的时序逻辑规划。《IEEE 机器人学汇刊》，2022年。[Kasenberg 等， 2020] Daniel Kasenberg、Ravenna Thielstrom 和 Matthias Scheutz。为时序逻辑规划器决策生成解释。载于《国际自动规划与调度会议论文集》，第 30 卷，第 449–458 页，2020年。[Lacerda 等， 2015] Bruno Lacerda、David Parker 和 Nick Hawes。面向部分可满足 co-safe LTL 规范的最优策略生成。载于第二十四届国际人工智能联合会议，2015年。[Lahijanian 和 Kwiatkowska， 2016] Morteza Lahijanian 和 Marta Kwiatkowska。具有最优权衡的马尔可夫决策过程规范修订。载于第 55 届决策与控制会议（CDC' 16）论文集，第 7411–7418 页，2016年。[Li 等， 2020] Meilun Li、Andrea Turrini、Ernst Moritz Hahn、Zhikun She 和 Lijun Zhang。马尔可夫决策过程的概率偏好规划问题。《IEEE 软件工程汇刊》，2020年。[Li 等， 2021] Shen Li、Daehyung Park、Yoonchang Sung、Julie A Shah 和 Nicholas Roy。时序逻辑规范下的反应式任务与运动规划。载于 2021年 IEEE 国际机器人与自动化会议（ICRA），第 12618–12624 页。IEEE，2021年。[Mallett 等， 2021] Ian Mallett、Sylvie Thi´ebaux 和 Felipe Trevizan。带有概率 LTL 约束的规划进展启发式。载于 AAAI 人工智能会议论文集，第 35 卷，第 11870–11879 页，2021年。[Mehdipour 等， 2021] Noushin Mehdipour、Cristian-Ioan Vasile 和 Calin Belta。使用加权信号时序逻辑表达用户偏好。《IEEE 控制系统快报》，5(6):2006–2011，2021年12月。[Pnueli， 1981] Amir Pnueli。并发程序的时序语义。《理论计算机科学》，13(1):45–60，1981年。[Puterman， 2014] Martin L Puterman。马尔可夫决策过程：离散随机动态规划。John Wiley & Sons，2014年。

[Rahmani 和 O’Kane， 2019] Hazhar Rahmani 和 Jason M O’Kane。具有级联软约束的最优时序逻辑规划。载于 *2019*年 *IEEE/RSJ* 智能机器人与系统国际会议（*IROS*），第 2524–2531 页。IEEE，2019年。[Rahmani 和 O’Kane， 2020] Hazhar Rahmani 和 Jason M O’Kane。无法面面俱到时该怎么办：带软时序逻辑约束的时序逻辑规划。载于*2020 IEEE/RSJ*智能机器人与系统国际会议（*IROS*），第6619–6626页。IEEE，2020年。

[Tumova等，2013a] Jana Tumova、Luis I Reyes Castro、Sertac Karaman、Emilio Frazzoli 和 Daniela Rus。针对相互冲突规范的最小违反线性时序逻辑规划。载于美国控制会议，第200–205页。IEEE，2013。

[Tumova等，2013b] Jana Tumova、Gavin C Hall、Sertac Karaman、Emilio Frazzoli 和 Daniela Rus。带安全规则的最少违反控制策略综合。载于混合系统：计算与控制国际会议论文集，2013。[Vasile等， 2017] Cristian-Ioan Vasile、Jana Tumova、Sertac Karaman、Calin Belta 和 Daniela Rus。面向按需出行的最小违反SCLTL运动规划。第1481–1488页。IEEE，2017。[Vasile等， 2020] Cristian Ioan Vasile、Xiao Li 和 Calin Belta。带时序逻辑规范的反应式采样路径规划。《国际机器人研究期刊》，39(8):1002–1028，2020年。[Wang等， 2020] Yu Wang、Siddhartha Nalluri 和 Miroslav Pajic。机器人学中的超性质：基于HyperLTL的规划。载于*2020 IEEE*机器人与自动化国际会议（*ICRA*），第8462–8468页。IEEE，2020年。[Wolper， 2001] Pierre Wolper。从时序逻辑公式构造自动机：教程。载于G. Goos、J. Hartmanis、J. van Leeuwen、Ed Brinksma、Holger Hermanns 和 Joost-Pieter Katoen（编），形式化方法与性能分析讲座，第2090卷，第261–277页。Springer Berlin Heidelberg，柏林、海德堡，2001年。[Wongpiromsarn 等， 2021] Tichakorn Wongpiromsarn、Konstantin Slutsky、Emilio Frazzoli 和 Ufuk Topcu。自主系统的最小违反规划：理论与实践考量。载于*2021*年美国控制会议，2021年。已投稿。[Yang 等， 2020] Yuanjiang Yang、Xiang Yin 和 Shaoyuan Li。面向多机器人任务规划的分布式框架，带有时序逻辑规范。载于*2020 IEEE*第*16*届控制与自动化国际会议（*ICCA*），第570–575页。IEEE，2020年。[Zhao 等， 2022] Jiawei Zhao、Xiang Yin 和 Shaoyuan Li。通过主动获取信息进行时序逻辑机器人任务规划。载于*2022 IEEE*第*17*届控制与自动化国际会议（*ICCA*），第1014–1020页。IEEE，2022年。

[Zhou 等， 2022] Xiaoyi Zhou、Tiange Yang、Yuanyuan Zou、Shaoyuan Li 和 Hao Fang。冲突信号时序逻辑任务下多智能体系统的多个子公式协同控制。*IEEE* 工业电子学汇刊，2022 年。

A 证明

引理 2 的证明

证明。 </style}运行 ρ 当且仅当 $q_{n-1} \in F$、$\sigma_{n-1} = \ltimes$ 和 $q_n =$ 汇点时具有权重 1。由于确定有限自动机的接受条件，词 Word(ρ) $= \sigma_0\sigma_1 \ldots \sigma_{n-1}$ 被接受，因而满足线性时序逻辑$_f$ 公式 φ.

□

引理 3 的证明

证明。给定倒数第二个状态 $\mathrm{q}_{n-1} = (q_1, q_2)$，如果 $\mathrm{w}_1(q_1, \ltimes,$ 汇点) $= \mathrm{w}_2(q_2, \ltimes,$ 汇点) $= 0$，那么 Word(ρ) $\not\models \varphi$，这意味着它不满足 φ。如果 $\mathrm{w}_1(q_1, \ltimes,$ 汇点) $= 0$，但 $\mathrm{w}_2(q_2, \ltimes,$ 汇点) > 0，那么 Word(ρ)在正的程度上满足 φ_2，但不满足 φ_1。相对于$\rightarrow\varphi_1 \times \varphi_2$ 的满足度是相对于 φ_2 的满足度与 φ_1的可选性之和。否则，如果 $\mathrm{w}_1(q_1, \ltimes,$ 汇点) > 0，那么相对于 $\varphi_1 \times \varphi_2$ 的$\overrightarrow{\text{满足度就是相对于}}$ φ_1的满足度。

□

引理 4 的证明

证明。 该证明通过构造给出，与引理 3的证明类似。因此，省略证明。 □

引理 5 的证明

证明。 该性质可基于递归定义得到证明。首先，显然，如果 PQCLTL$_f$ 公式是一个线性时序逻辑$_f$ 公式，则可选性为 1，且所有已定义转移的最大权重为 1。考虑两个 PQCLTL$_f$ 公式 φ_1、φ_2及其对应的加权确定有限自动机 $\mathcal{A} = \langle Q_i \cup \{$汇点$\}$、$\Sigma \cup \{\ltimes\}$、$\delta_i$、$q_{0i}$，$\mathbf{w}_i\rangle$满足$opt(\varphi_i) =$最大值$\{\mathbf{w}_i(q,$ a、 $q') \mid \delta_i(q,$ a、 $q')$ 已定义。$\}$。

$\overrightarrow{\text{在有序析取}}$ $\varphi_1 \times \varphi_2$ 的加权确定性有限自动机中，按照构造，最大权重为 $\max\{\mathrm{w}_2(q_2, \ltimes,$汇点) $+opt(\varphi_1)\} = \max\{\mathrm{w}_2(q_2, \ltimes,$ 汇点)$\} + opt(\varphi_1) = opt(\varphi_2) + opt(\varphi_1)$，这与定义 4 一致。

在优先级合取 φ_1&φ_2 的 WDFA 中，根据构造，最大权重为 $\max_{i,j}\{opt(\varphi_2) \times (i-1) + j\}$ 其中 $0 < i \le \max\{\mathrm{w}_1(q_1, \ltimes,$ 汇点$)\} = opt(\varphi_1)$ 以及 $0 < j \le \max\{\mathrm{w}_2(q_2, \ltimes,$ 汇点$)\} = opt(\varphi_2)$。因此，$\max_{i,j}\{opt(\varphi_2) \times (i-1) + j\} = opt(\varphi_2) \times (opt(\varphi_1) - 1) + opt(\varphi_1) = opt(\varphi_2) \times opt(\varphi_1)$。这再次与定义 4 一致。

□

引理 6 的证明。

证明。 有限运行 $\rho = v_0a_0v_1a_1 \ldots v_n$ 只有在存在 $0 \le k \le n$, $v_k =$ ($s_\perp$，汇点) 且对于所有 $j \le k, v_j \neq$ ($s_\perp$，汇点) 时才会获得非零奖励。 ρ 的总奖励上界为 opt (φ)。因此，对于任意策略 π 和任意状态 v，$J_{\pi,\ N}(v)$ 在 $N \to \infty$ 存在时的极限存在，且其上界为 $opt(\varphi)$。 □

引理 7 的证明。

证明。 我们证明，对于每个不正规的最优马尔可夫策略，都存在一个可获得相同价值的正规马尔可夫策略。考虑一个不正规的最优马尔可夫策略 $\pi^\dagger$，在该策略下存在一条无限运行。由于只有到达状态 ($s_\perp$，汇点) 时才能获得奖励，因此该无限运行 h的奖励为零。因此，可以通过对所有有限运行复制 $\pi^\dagger$，构造一个具有相同价值 $J_{\pi^\dagger}(v) = J_{\pi^*}(v)$ 的正规策略 π^*。对于所有无限运行，可通过在任意奖励为零的状态终止，从 $\pi^\dagger$ 得到 π^*。

□

引理 8 的证明。

证明。通过构造进行证明。对于第一种情况，回顾$R(h) = \sum_i^{n-1} R(v_i, a_i)$。如果对于某个 $k > 0$， $L(\mathrm{Proj}_S(h)) \models_k \varphi$，那么 $R(h) = opt(\varphi) - \mathrm{w}(q_n, \ltimes,$ 汇点$) + 1$。代入$R(h)$，可得 $\mathrm{d}(L(\mathrm{Proj}_S(h)) \ \varphi) = 1 - {}^{R(h)}{}_{opt(\varphi)+1} = \frac{\mathrm{w}(q_n, \ltimes, \text{汇点})}{opt(\varphi)+1}$，~~符合~~引理 2、3 和 4。对于第二种情况，如果 $L(\mathrm{Proj}_S(h)) \not\models \varphi$，那么 $R(h = 0)$，于是 $\mathrm{d}(L(\mathrm{Proj}_S(h)) = 1$，符合定义 6。 □

B 加权自动机构造示例

我们通过一个示例说明加权确定性有限自动机的构造。[{"id":0,"output":"示例 1。给定两个 LTL$_f$ 公式 $\Diamond b$ 和 $\Diamond a \vee \Diamond c$，以及一个 PQCLTL$_f$ 公式 $\Diamond b \overrightarrow{\times} (\Diamond a \vee \Diamond c)$，其含义是“如果可能，最终满足 b；如果不可能，则最终满足 a 或c。”LTL$_f$ 公式对应的 WDFA 如图 4a 和 4b 所示，WDFA 如图 4c 所示。为清晰起见，我们使用命题逻辑公式而不是 $2^{\mathcal{AP}}$ 作为转移的符号。例如， $b \wedge (a \vee c) : 0$ 表示$\{b, a\} : 0$、 $\{b, c\} : 0$ 和 $\{b, a, c\} : 0$。从图 4c 可以看出，从 (0, 1) 到汇点的权重转移为 2，这是因为触发该转移后，满足的公式 $\Diamond a \vee \Diamond c$的优先级较低。"}]

在表1 中，我们列出了不同词语对应的满足度。根据不满意度分数，我们有 $w_1 \succeq^\phi w_2 \succeq^\phi w_3$，其中 $\phi = \Diamond b \overrightarrow{\times} (\Diamond a \vee \Diamond c)$。

表1：词语的不满意度分数 相对于 $\Diamond b \overrightarrow{\times} (\Diamond a \vee \Diamond c)$

词语	词语描述	$w \models_i \Diamond b$	$w \models_j \Diamond a \vee \Diamond c$	$w \models_k \Diamond b \overrightarrow{\times} (\Diamond a \vee \Diamond c)$	$\mathrm{d}(w, \Diamond b \overrightarrow{\times} (\Diamond a \vee \Diamond c))$
w_1	$\{b\}\{a\}\ltimes$	1	1	1	$1/3$
w_2	$\emptyset\emptyset\{a\}\ltimes$	$\not\models$	1	2	$2/3$
w_3	$\emptyset\emptyset\ltimes$	$\not\models$	$\not\models$	$\not\models$	1

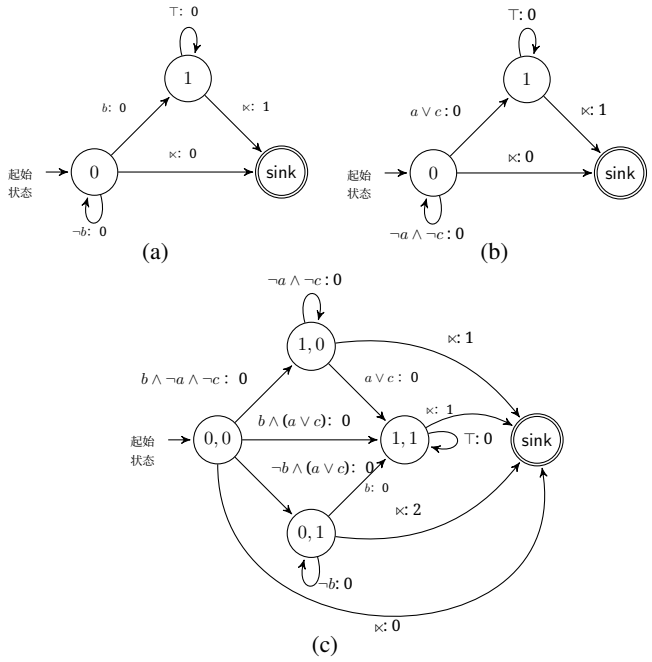


图4：（a）接受公式 $\Diamond b$ 的加权确定性有限自动机。（b）接受公式 $\Diamond a \vee \Diamond c$ 的加权确定性有限自动机。（c）接受公式 $\Diamond b \overrightarrow{\times} (\Diamond a \vee \Diamond c)$ 的加权确定性有限自动机。