arXiv:2603.05508v1 [cs.FL] 14 Jan 2026

离散事件系统的标记数据信息性与数据驱动监督控制

Yingying Liu[1], Kuma Fuchiwaki[1], 和 Kai Cai[1]

摘要——本文提出一种用于离散事件系统（DES）标记非阻塞监督控制的数据驱动方法。我们考虑这样一种情形：待控离散事件系统的模型未知，但可获得一组关于离散事件系统行为的数据。我们探讨以下问题：在什么条件下，利用现有数据集能够为未知离散事件系统设计有效的标记非阻塞监督器，以满足给定规范？为回答这一问题，我们提出并形式化定义了一个称为标记数据信息性的新概念。此外，我们设计了一种用于验证该概念的算法。接下来，如果数据集不具备标记信息充分性，我们将提出两个相关的新概念：受限标记数据信息性和标记可信息化性。最后，我们开发了一种算法，用于计算使数据集具有最小受限标记信息性的最大控制规范子集。

I. 引言

状态因事件发生而离散转移的系统称为离散事件系统（DES）。交通系统、生产系统和通信系统都是离散事件系统的例子 [1]–[2]。监督控制被视为离散事件系统的一种有效控制方法。为了使给定离散事件系统满足期望的规范，监督控制理论旨在构造一个称为监督器的反馈控制器，其控制机制是禁用或启用某些（可控）事件的发生。自动驾驶和配送仓库自动化是监督控制的典型应用 [3], [4], [5], [6], [7], [8]。

近年来，随着物联网的普及和数据感知技术的快速发展，数据驱动方法在各个领域备受关注 [9]–[10]。这意味着许多系统的可用数据正在迅速增长 [11], [12]。另一方面，离散事件系统的监督控制大多采用基于模型的方法，鲜有研究尝试采用数据驱动方法 [13], [14]–[15]。这些正是研究数据驱动监督控制的动机。

在基于模型的监督控制中，首先将待控离散事件系统建模为有限状态自动机，并用正则语言表示其行为，然后基于这些模型进行监督控制设计。另一方面，我们的数据驱动方法并不旨在辨识离散事件系统模型，而是处理与离散事件系统行为有关的数据集。其目标是为能够生成该数据集的离散事件系统模型族构造一个有效监督器。这些模型都能够生成该数据集。这使得我们能够有效控制环境未知且难以建模的离散事件系统。先前研究提出了数据信息性的概念，用以刻画这样一种观念：可用数据集包含足够的信息，从而能够为能够生成该数据集的离散事件系统模型族构造有效监督器；同时还提出了数据信息性的充要条件及其验证算法 [16], [17]。然而，先前研究的表述未考虑标记行为。本文通过纳入考虑标记行为的数据信息性，弥补了这一不足。此方法改进了监督控制设计，使系统能够避免死锁，同时确保实现指定目标 [18]。

在本论文中，我们旨在系统地发展一种用于离散事件系统标记监督控制的数据驱动方法，并通过仿真实例验证其有效性。具体而言，我们考虑这样一种设置：待控制的离散事件系统模型是未知的，但可以获得关于离散事件系统行为的三类数据。第一类是观测数据 D，即来自未知离散事件系统的观测行为（事件字符串）的集合。第二类是关于标记行为的观测数据 D_m。第三类是关于离散事件系统不可能行为的先验知识数据 D^- 。给定这三类数据，我们研究这样一个问题：对于可用数据集，需要满足什么条件，才能为未知离散事件系统设计出符合给定规范的有效标记非阻塞监督器。通过回答这一问题，我们提出并形式化了一种称为标记数据信息性的新概念。标记数据信息性刻画了这样一种条件：给定数据集所包含的信息足以使得能够为一族均可生成该数据集的离散事件系统模型构造有效的标记非阻塞监督器。因此，我们的方法并不试图首先识别未知离散事件系统的模型，而是基于标记数据信息性，旨在直接从数据集中构造一个对所有与未知离散事件系统无法区分的可能模型均有效的标记非阻塞监督器。如果能够构造出这样的标记非阻塞监督器，那么它对未知离散事件系统也同样有效。

论文的结构如下：第二节回顾预备知识。第三节介绍数据驱动的标记监督控制和标记数据信息性，

[1]大阪公立大学核心信息学系，5588585 大阪，日本 cai@omu.ac.jp

而第四节讨论受限标记数据信息性。第六节总结全文。

II. 预备知识与问题陈述

设 Σ 为由符号 σ、α、β.....组成的非空有限字母表。这些符号表示事件， Σ 为事件集。字符串 $s = \sigma_1\sigma_2\ldots\sigma_k$, $k \leq l$, 是事件的有限序列。设 Σ^* 为包含空串 ε（表示“不执行任何操作”）在内的所有有限长度字符串的集合。语言 L 是 Σ^* 中字符串的任意子集，即 $L \subseteq \Sigma^*$。如果存在事件序列 $s_2 \in \Sigma^*$，使得 $s = s_1s_2$,，则事件序列 $s_1 \in \Sigma^*$ 称为 s 的前缀。对于任意语言 $L \subseteq \Sigma^*$，其元素的所有前缀所构成的语言可表示为 $\overline{L} := \{s_1 \in \Sigma^* | (\exists s_2 \in \Sigma^*) s_1s_2 \in L\}$.。恒有 $L \subseteq \overline{L}$。若语言 L 满足 $L = \overline{L}$，则称其为闭的。

自动机是离散事件系统的基础模型。有限状态自动机 G 是一个五元组

$$G := (Q, \Sigma, \delta, q_0, Q_m) \tag{1}$$

其中，Q 是有限状态集，$q_0 \in Q$ 是初始状态，$Q_m \subseteq Q$ 是标记状态集， Σ 是有限事件集，而 $\delta : Q \times \Sigma \to Q$ 是（部分）状态转移函数。记 δ(q σ)!表示 $\sigma \in \Sigma$ 在状态 $q \in Q$ 处有定义，并记¬δ(q σ)!表示 $\sigma \in \Sigma$ 在 $q \in Q$ 处未定义。若自动机 G 满足下式，则称其为确定性的

$$(\forall q \in Q, \forall \sigma \in \Sigma)\ \ \delta(q, \sigma)! \Longrightarrow |\delta(q, \sigma)| = 1.$$

也就是说，每个状态转移的目标状态都是唯一的。除非另有说明，本文将只讨论确定性自动机。状态转移函数 δ可以通过归纳定义，我们记δ(q,s)!表示 δ(q, s)有定义。G的闭行为L(G)是L(G) := {s ∈ Σ*|δ(q0,s)!}.。根据定义，L(G)是闭的。此外，G的标记行为L_m(G)是如下定义的语言：由G从初始状态q_0出发到标记状态Q_m所能生成的L(G)中的所有字符串构成，定义为L_m(G) := {s ∈L(G)|δ(q0,s) ∈ Qm}.。与自动机的定义类似，若(∃s ∈Σ*)δ(q0,s)! & δ(q0, s) =q，则称状态q ∈Q是可达的。如果对所有q ∈Q，q均可达，则称G本身是可达的。若(∃s ∈ Σ*)δ(q, s) ∈ Qm·，则称状态q ∈ Q是共可达的。如果对所有q ∈ Q，q均共可达，则称G本身是共可达的。若每个可达状态都是共可达的，则称G是非阻塞的；等价地，L(G) = L_m(G)，也就是说，闭行为中的每个字符串都可以扩展为标记行为中的一个字符串。最后，若G既可达又共可达，则称其为整洁的。显然，整洁蕴含非阻塞，但反之不成立。

标准的基于模型的监督控制由三个要素组成：被控对象、控制规范和监督器（控制器）。被控对象用确定有限自动机 G 建模。监督器会观察被控对象中发生的事件，并禁止不适当的事件，以确保被控对象满足控制规范。

通常，被控对象 G 的标记行为 $L_m(G)$ 中并非所有字符串都是期望的。因此，我们将要在 G 上强制实现的期望行为表示为控制规范$K \subseteq L_m(G)$。请注意，也可以考虑更一般的控制规范$E \subseteq \Sigma^*$。在这种情况下，令 $K := L_m(G) \cap E$，于是我们仍得到一个要在 G 上强制实现的规范 $K \subseteq L_m(G)$。

为了在被控对象 $K \subseteq L_m(G)$ 上实施规范，我们假设事件$\Sigma_c \subseteq \Sigma$的一个子集（称为可控事件）可以由外部控制器启用或禁用。相反， $\Sigma_u := \Sigma \setminus \Sigma_c$ 是不可控事件的集合，这些事件无法由外部禁用，必须始终视为已启用。

在上述机制下，将监督器定义为函数 $V : L(G) \to Pwr(\Sigma_c)$。此处，$Pwr(\Sigma_c)$ 表示可控事件所有子集的集合。因此，对于被控对象 G 生成的每个字符串 $s \in L(G)$，监督器都会指定一个要禁用的可控事件子集 $V(s) \subseteq \Sigma_c$。用 V/G 表示闭环系统。闭环系统的语言 $L(V/G)$ 定义如下：

1) $\varepsilon \in L(V/G)$;
2) if $s \in L(V/G)$ and $\sigma \in \Sigma \setminus V(s)$ and $s\sigma \in L(G)$, then $s\sigma \in L(V/G)$;
3) no other strings belong to $L(V/G)$.

因此，$L(V/G)$ 包含 $L(G)$ 中未被监督器 V 禁用的字符串。根据定义，$\overline{L(V/G)} \subseteq L(G)$ 且 $L(V/G) = L(V/G)$。

此外，闭环系统 $L_m(V/G)$ 的标记语言定义如下：

$$L_m(V/G) = L(V/G) \cap L_m(G).$$

若闭环系统的标记语言为，则称监督器V 为标记监督器

$$L_m(V/G) = L(V/G) \cap K.$$

即监督器V 负责规范 K 的标记信息。此外，若 $L_m(V/G) = L(V/G)$，则称V 为标记非阻塞监督器。

定义1：（可控性）：给定被控对象 G，控制规范 $K \subseteq L_m(G)$ 被称为关于 G 可控，如果

$$(\forall s \in \overline{K}, \forall \sigma \in \Sigma_u)\ s\sigma \in L(G) \Rightarrow s\sigma \in \overline{K}. \tag{2}$$

换言之，规范 K 关于 G 是可控的，当且仅当前缀闭包 K 中的任一字符串都不能通过一个不可控事件的延续而离开 K。也就是说，K 的前缀闭包在不可控流下是不变的。等价地，我们可以写成 $K\Sigma_u \cap L(G) \subseteq K$。已知，规范语言$K(\neq\emptyset)$ 可控是存在标记非阻塞监督器V 使得 $L_m(V/G) = K$ 且 $L_m(V/G) = L(V/G)$ 的充要条件。

假设 $K \subseteq L_m(G)$ 是可控的。那么标记非阻塞监督器 V: $L(G) \to Pwr(\Sigma_c)$ 满足

如下构造 $L_m(V/G) = K$：

$$V(s) = \begin{cases} \{\sigma \in \Sigma_c \mid s\sigma \notin K\} & \text{if } s \in \overline{K}, \\ \emptyset & \text{if } s \in L(G) \setminus \overline{K}. \end{cases} \quad (3)$$

无论K 是否可控，我们都可以将$C(K)$表示为K的所有可控子语言的族：

$$C(K) := \{K' \subseteq K \mid \overline{K'}\Sigma_u \cap L(G) \subseteq \overline{K'}\}. \quad (4)$$

已知K 的可控子语言的并仍然是K的可控子语言。这意味着$C(K)$在集合并运算下是闭的，因此$C(K)$包含唯一的上确界元素：

$$\sup C(K) := \bigcup \{K' \mid K' \in C(K)\}. \quad (5)$$

由于$\sup C(K)$是可控的，只要$\sup C(K) \neq /0$，就存在一个标记非阻塞监督器$V_{\sup}$，使得$L_m(V_{\sup}/G) = \sup C(K)$。从这个意义上说，$V_{\sup}$是最优的（最大许可的），允许$G$ 生成满足给定规范的最大可能字符串集。

例 1：为便于说明，我们给出一个机器人导航的贯穿示例。考虑一个从起点移动到目标点的机器人。沿途有一些路径会通向危险区域，机器人在朝目标前进时必须避开这些路径。此外，机器人还可能因环境中的潜在扰动而在某些位置发生不可控移动。图1展示了对此情景建模的自动机，我们将此自动机（记作 G_1）视为待控的被控对象。在该被控对象中，以数字表示的每个状态都代表机器人导航所在环境中的一个位置。这里，状态 0 是起点，状态 7 是目标点。这个目标状态是标记

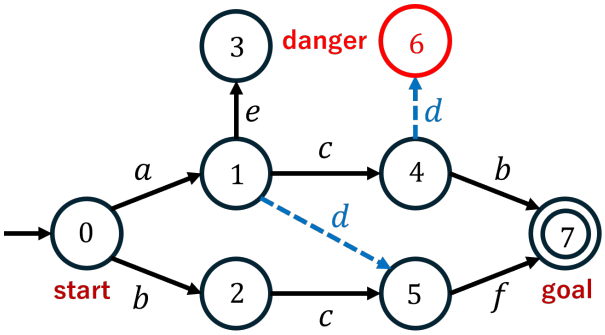


图1。机器人导航：被控对象G_1。

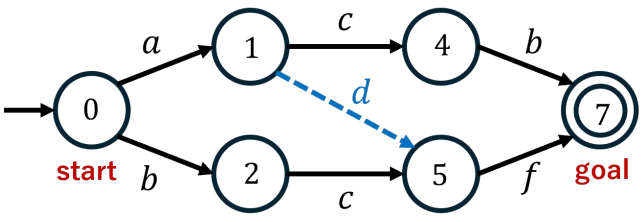


图2。机器人导航：规范$K_1(\subseteq L_m(G_1))$。

状态，以双圆圈表示。字母表中的每个事件代表机器人的一次转移。我们假设事件 d （从状态 1 到状态 5、从状态 4 到状态 6 的虚线箭头）表示不可控转移，其他事件均可控：即 $\Sigma_c = \{a, b, c, e, f\}$以及 $\Sigma_u = \{d\}$。状态 6 表示危险状态，控制规范是避开该危险状态。此（安全）规范可写为 $L_m(G_1)$ 的一个子语言，如下所示：

$$K_1 = \{acb, adf, bcf\}.$$

该规范K_1 可以由图2所示的自动机表示。与图1中的被控对象相比，该规范自动机移除了从状态4到危险状态6的（不可控）转移，以及从状态1到状态3的（可控）转移。

由于不可控事件 d 可以离开 K_1，，规范语言 K_1 是不可控的。接着，考虑 K_1 的子语言 $K_1' = \{$ad f ,bc f $\}$（从 K_1 中删除路径 acb）。在这种情况下，没有不可控事件可以离开 K_1'，因此 K_1' 是可控的，且 $\sup C(K_1) = K_1'$。因此，我们可以构造如下标记非阻塞监督器 V_1： $L(G_1) \to Pwr(\Sigma_c)$，使得 $L_m(V_1/G_1) = K_1'$：

$$V_1(s) = \begin{cases} \{c, e\} & \text{if } s \in \{a\}, \\ \emptyset & \text{if } s \in L(G_1) \setminus \{a\}. \end{cases}$$

该监督器在图1的状态1处禁用 c 和e 。

在上述示例中，监督器V_1 是在假设被控对象模型 G_1 已知的前提下设计的。现在我们提出这样一个问题：如果G_1 未知（例如，机器人在未知环境中导航），在什么条件下我们仍然能够设计标记非阻塞监督器？这个问题促使我们研究一种用于标记非阻塞监督控制的数据驱动方法。

III. 数据驱动标记监督控制与标记数据信息性

第3节在III-A节中表述了数据驱动标记监督控制问题。在第III-B节中，我们引入标记数据信息性的概念，并建立了刻画该概念的充要条件。基于此条件，第III-C节提出了一种用于验证标记数据信息性的算法。

A. 数据驱动标记监督控制的问题表述

假设我们有一个被控对象，其自动机模型G 是未知的，只有事件集 $\Sigma(=\Sigma_c \cup \Sigma_u)$ 已知。即使在这种情况下，也常常可以观测到被控对象生成的字符串，也就是说，可以获得一定量的输出序列数据。此外，通常还可以从观测中识别出到达标记状态的字符串。另外，根据事件集 Σ 的先验知识，往往还能确定某些显然不

可能由被控对象生成的输出序列。例如，假设有一组事件：启动机器、按下机器上的开关，以及机器产生输出。那么，即使不了解机器的内部工作机制，也显然可以知道，机器在通电前不可能输出任何内容。鉴于此，我们假设可以获得三类有限数据集 (D, D_m, D^-)，其中 $D \subseteq \Sigma^*$ 是被控对象的观测行为，$D_m \subseteq D$ 是观测到的标记行为的子集，而 $D^- \subseteq \Sigma^*$ 是关于被控对象不可能行为的先验知识。由于 D 和 D_m 中的每个字符串都是从 G 观测到的，因此 D 是 G 的闭行为的子集，且 D_m 是 G 的标记行为的子集：即 $D \subseteq L(G)$，$D_m \subseteq L_m(G)$。相反，由于已知 D^- 中的每个字符串都不可能由 G 生成，因此 D^- 与 L(G) 的闭行为没有任何共同元素：即 $D^- \cap L(G) = /0$。因此，$D \cap D^- = /0$。给定控制规范 $E \subseteq \Sigma^*$，我们的目标是基于数据集 (D, D_m, D^-)，为未知被控对象设计一个标记非阻塞监督器（若其存在），以强制实现 E。

例 2：再次考虑例 1 中的机器人导航示例以及图1中的被控对象模型 G_1。现在，假设除了事件集 $\Sigma = \{a, b, c, d, e, f\}$ 之外，G_1 未知，但我们可以获得被控对象行为的某些观测和先验知识。例如，我们观测到被控对象生成了字符串 bcf，即机器人沿路径 bcf 从初始状态移动到目标状态。因此，$D = \{bcf\}$ 且 $D_m = \{bcf\}$。此外，我们有先验知识，知道被控对象无法生成字符串 e，也就是说，机器人绝不会从位置 1 开始导航并首先移动到位置 3。因此，$D^- = \{e\}$。对于这个三元组 (D, D_m, D^-)，可能存在无穷多个自动机，它们能够生成 D 并标记 D_m，但不会生成 D^-。换言之，基于三元组 (D, D_m, D^-)，无法唯一识别未知被控对象 G_1。例如，图3中的 G_2 和图4中的 G_3 都无法与真实被控对象 G_1区分。为了仅基于 (D, D_m, D^-) 为真实被控对象设计监督器，我们必须构造一个对所有这类可能的被控对象都有效的监督器。直观而言，更多的观测和先验知识有助于减少那些无法与真实被控对象区分的被控对象模型的数量。比如，如果我们额外观测到字符串 adf（使得 $D = \{adf, bcf\}$，$D_m = \{adf, bcf\}$），那么 G_3 就可以从候选对象中排除，而 G_2 仍然是可能的。

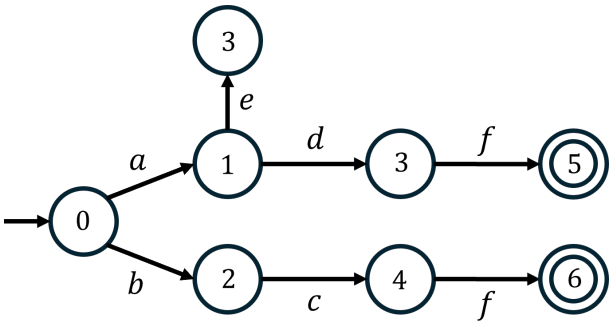


图3。示例2中的G_2。

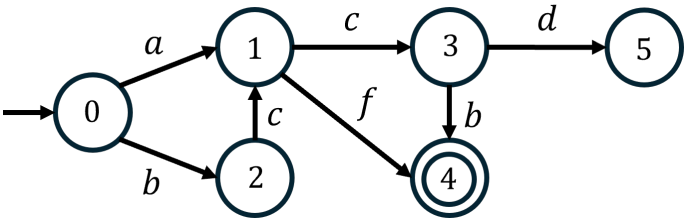


图4。示例2中的G_3。

如上例所示，与给定数据三元组 (D, D_m, D^-) 相容的被控对象通常不止一个。下文将其定义为一致性性质。

定义2：（一致性）。设给定事件集 Σ。对于满足 $D_m \subseteq D$且$D \cap \overline{D^-} = /0$的有限集D、$D_m$、$D^- \subseteq \Sigma^*$，若自动机 $G = (Q, \Sigma, \delta, q_0, Q_m)$满足$D \subseteq L(G)$、$D_m \subseteq L_m(G)$以及$D^- \cap L(G) = /0$，则称G与$(D, D_m, D^-)$一致。

换言之，当且仅当D中的所有字符串都可由G生成、D_m 中的所有字符串都可由G标记，而D^- 中的字符串均不可由G生成时，自动机G才与数据三元组(D, D_m, D^-)一致。因此，在示例2中，G_1、G_2和G_3与(D, D_m, D^-)一致，其中$D = \{bcf\}$、$D_m = \{bcf\}$且$D^- = \{e\}$。如果我们额外观测到字符串adf（$D = \{adf、bcf\}$、$D_m = \{adf、bcf\}$且$D^- = \{e\}$），G_1和G_2仍与(D, D_m, D^-)一致，但G_3 将变得不一致。容易验证，如果我们观测到一些额外字符串，或对被控对象无法生成的字符串有更多了解，一致模型的数量就会减少。

注1：众所周知 [19]，对于每一个正则语言$L_m \subseteq \Sigma^*$，都存在一个自动机G，使得$L_m(G) = L_m$。因此，对于任意满足$D_m \subseteq D$且$D \cap \overline{D^-} = /0$的有限集D、$D_m$、$\overline{D^-} \subseteq \Sigma^*$（它们是正则的），总能找到一个自动机G，使得$L(G) = D$、$L_m(G) = D_m$，从而$D^- \cap L(G) = /0$。这意味着至少存在一个与$(D, D_m, D^-)$一致的自动机G。

在继续之前，我们总结一致对象的一些基本性质。

命题1：存在一个与（D，D_m，D^-）一致的被控对象，当且仅当$D_m \subseteq D$是正则的，且$D \cap \overline{D^-} =$

/0. 如果G与（D，D_m，D^-）一致，那么G也与（D，D_m，D^-）一致。此外，$\overline{\ }$有

$$\bigcap\{L(G)\mid G \text{ is consistent with } (D,D_m,D^-)\}=\overline{D}. \quad (6)$$

命题1的证明可由定义2和备注1直接得出。

命题1的最后一个断言（6）促使我们研究对D的监督控制。现在，我们表述数据驱动标记监督控制问题。我们将基于D_m的控制规范记为：

$$K_{D_m} := D_m \cap E, \quad E \subseteq \Sigma^* \text{ (regular language)}. \quad (7)$$

定义3：给定数据集（D，D_m，D^-），设G是与（D，D_m，D^-）一致的被控对象，且V_D：$D \to Pwr(\overline{\Sigma_c})$是$D$的标记非阻塞监督器。我们将$V_D/G$ 基于D_m 的标记语言定义为

$$L_m(V_D/G) = L(V_D/G) \cap D_m. \quad (8)$$

问题1：设给定事件集$\Sigma = \Sigma_c \cup \Sigma_u$、控制规范E $\subseteq \Sigma^*$以及有限数据集D、D_m、$D^- \subseteq \Sigma^*$，使得（7）中的K_{D_m}非空、$D_m \subseteq D$，且$D \cap D^- = /0$。构造（若可能）一个标记非阻塞监督器V_D：$D \to Pwr(\Sigma_c)$，使得对于每个与（D、D_m、D^-）一致的被控对象G，都有$L_m(V_D/G) = K_{D_m}$。

由于真实被控对象与 (D,D_m, D^-) 一致，因此，满足问题1中所要求条件的监督器对真实被控对象有效。如果已知真实被控对象 G，我们就会构造一个标记非阻塞监督器，以实现 K $= L_m(G) \cap E$（只要 K 可控）。在我们的数据驱动设置中，D 和 D_m 分别表示基于对被控对象的观测所得到的 L(G) 和 $L_m(G)$ 的可访问子集；因此，根据现有数据，最多只能构造一个监督器来实现 $K_{D_m} = D_m \cap E$。如果 K_{D_m} 所表示的行为过小或限制过多，可以通过观测被控对象的更多行为来考虑扩大 D 和 D_m。D 和 D_m 分别越接近 L(G) 和 $L_m(G)$，数据驱动的可实现行为 K_{D_m} 就越接近原始的基于模型的行为 $L_m(G) \cap E$。

B. 标记数据信息性及其标准

如 III-A节所述，存在一个标记非阻塞监督器 V_D，使得对于所有与（D，D_m，D^-）相一致的被控对象 G，均有 K_{D_m}（$\neq /0$）$\subseteq G$）$= K_{D_m}$。这等价于规范语言 K_{D_m}（$\neq /0$）的可控性。这表明，可用数据是否包含充分信息，可以用可控性来刻画。

定义4：（标记数据信息性）。对于给定的控制规范 D，D_m，D^-，如果存在满足问题1中所要求条件的标记非阻塞监督器，或者等价地，如果式（7）中的 K_{D_m} 非空，且相对于所有与 (D，D_m，D^-) 相一致的被控对象 G均可控，则称其具有标记信息充分性。

回顾一下，对于已知被控对象，如果某个不可控事件 $\sigma \in \Sigma_u$ 可以在被控对象中 $s \in K_{D_m}$ 之后发生，那么为了保证 K_{D_m} 相对于该被控对象的可控性，$s\sigma$ 必须保留在 K_{D_m} 中；参见式（2）。相反，对于数据驱动情形（不了解被控对象），除非$s\sigma \in D^-$（已知不可能），否则我们需要假设任意不可控事件 $\sigma \in \Sigma_u$ 都可以在 $s \in K_{D_m}$ 之后发生。由此得到以下观察：

定理1：（标记信息充分性判据）。设给定一个事件集 $\Sigma = \Sigma_c \cup \Sigma_u$ 和一个控制规范 $E \subseteq \Sigma^*$。（D，D_m，D^-）对于E 是标记信息充分的，当且仅当

$$(\forall s \in \overline{K_{D_m}}, \forall \sigma \in \Sigma_u) \quad s\sigma \in \overline{K_{D_m}} \cup D^- \quad (9)$$

其中K_{D_m} 见式（7）。

证明：（若）设式（9）成立。则容易验证

$$(\forall s \in \overline{K_{D_m}}, \forall \sigma \in \Sigma_u) \quad s\sigma \in L(G) \implies s\sigma \in \overline{K_{D_m}} \quad (10)$$

对每个与（D，D_m，D^-）相一致的被控对象 G均成立。因此，（D，D_m，D^-）对于E是标记信息充分的。

（仅当）假设 (D, D_m,D^-) 对 E 具有标记信息性。根据定义4，这意味着对于与 (D, D_m,D^-) 一致的每个被控对象 G，(10) 都成立。考虑一个特殊被控对象 G′，使得 L(G′) $= \Sigma^* \setminus D^-$。该 G′ 与 (D, D_m,D^-) 一致，并且任何不属于 L(G′) 的字符串都属于 D^-。令 s $\in K_{D_m}$ 且 $\sigma \in \Sigma_u$。分两种情况。情况1：$s\sigma \in$ L(G′)。由于 (10) 对 G′ 成立，因此 $s\sigma \in K_{D_m}$。情况2：$s\sigma \notin$ L(G′)。此时，$s\sigma \in D^-$。因此，(9) 得以满足。

定理1给出了标记数据信息性的一个充要条件。将(9)与K_{D_m}的标准可控性条件相比较，$L(G)$的缺失以及D^-的相应部分标示了数据驱动框架中的区别。

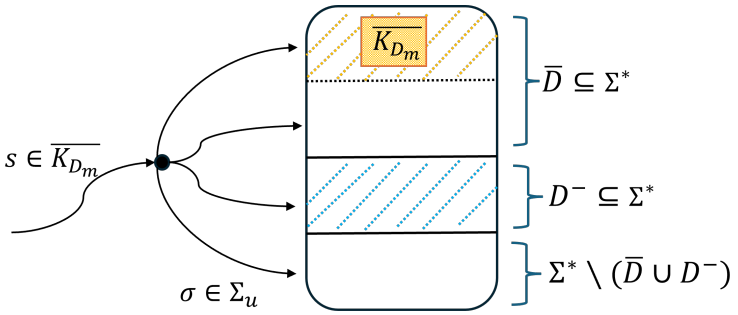


图5. (9) 中$s \in K_{D_m}$和 $\sigma \in \Sigma_u$ 的跃迁示意图。

为说明定理1，将 (3.4) 等价地写成以下两个方程：

$$(\forall s \in \overline{K_{D_m}}, \forall \sigma \in \Sigma_u) \quad s\sigma \in \overline{D} \implies s\sigma \in \overline{K_{D_m}} \quad (11)$$

$$\text{and} \quad s\sigma \notin \overline{D} \implies s\sigma \in D^-. \quad (12)$$

仅使用给定的数据集（D，D_m，D^-），要确定K_{D_m} 对于所有与其一致的被控对象G 是否可控，关键在于评估

字符串 s $\in K_{D_m}$ 的转移以及事件 $\sigma \in \Sigma_u$，如图 5 所示。(3.6)表示在观测所识别出的未知被控对象 G 中，K_{D_m} 在 D 内的可控性，对应于图 5 中的黄色区域。若 sσ ∈ D \ K_{D_m}，则说明在已观测范围内可控性已被破坏，(D, D_m, D^-) 不具备标记信息充分性。与式 (2) 中的基于模型的方法不同，后者规定 K_{D_m}中的任何字符串都不能通过不可控事件的延续而离开 K_{D_m}；数据驱动方法则要求考虑未观测字符串 sσ ∉D 的可能性。在这种情况下，必须确保此类 sσ不会出现在 G 中，即 sσ ∈ D^{-}。式 (12) 正好刻画了这一要求，对应于图 5 中的浅蓝色区域。最后，若 sσ ∈ $\Sigma^* \setminus (D \cup D^-)$，则给定数据集缺乏足够信息，因此不具备标记信息充分性。

下面我们将阐述D^- 在影响数据集质量方面所发挥的关键作用。

注2：集合D^- 包含无法由未知被控对象生成的字符串（即关于被控对象不可能行为的先验知识）。如果我们掌握的此类先验知识很少（即 $|D^-| \to 0$），那么为了满足 (9)，K_{D_m} 必须包含其自身字符串的几乎所有单步不可控延续。这反过来要求观测数据集D 和D_m 对于不可控事件的发生具有相当的完备性，而这在实践中可能难以实现（因此 (9) 难以满足）。因此，通常而言，更多的先验知识D^- 能够有效降低获取观测数据D 和D_m的要求。不过请注意，D^- 中的字符串并非都同样有用；根据 (9)，D^- 中有用的字符串，恰好是“K_{D_m}”中字符串的单步不可控延续。综上，对于标记信息性而言，重要的并非D、D_m 或D^-的单纯数量，而是 (9) 所规定的D、D_m 与D^- 之间的“匹配性”质量。

命题 2：假设给定一个有限数据三元组（D,D_m,D^-），它对给定规范 E 是标记信息充分的。则一个标记非阻塞监督器 V_D：$D \to Pwr(\Sigma_c)$，使得 $L_m(V_D/G) = K_{D_m}$，按如下方式构造：

$$V_D(s) = \begin{cases} \{\sigma \in \Sigma_c \mid s\sigma \notin \overline{K_{D_m}}\} & \text{if } s \in \overline{K_{D_m}}, \\ \emptyset & \text{if } s \in \overline{D} \setminus \overline{K_{D_m}}. \end{cases} \quad (13)$$

上述命题可直接借助模型已知情形中所用的标记非阻塞监督器构造方法导出（如（3）所示）。

例3：我们再次使用机器人导航示例来说明标记数据信息性的概念。假设图1中的被控对象G_1 未知，但事件集 $\Sigma = \Sigma_c \cup \Sigma_u$ 已知，其中 $\Sigma_c = \{a,b,c,e,f\}$ 和$\Sigma_u = \{d\}$。考虑两个有限数据三元组 (D_1, D_{m1}, D_1^-) 和 (D_2, D_{m2}, D_2^-)，其中

$$\begin{aligned} D_1 &= \{ae, adf, bcf\}, \\ D_{m1} &= \{adf, bcf\}, \\ D_1^- &= \{d, bd, aed, add, bcd, adfd, bcfd\}, \\ D_2 &= \{acb, acd, bcf\}, \\ D_{m2} &= \{acb, bcf\}, \\ D_2^- &= \{d, bd, bcd, acbd, bcfd\}. \end{aligned}$$

设 $E = \{ae, acb, adf, bcf\}$。则（7）中的控制规范分别为

$$\begin{aligned} K_{D_{m1}} &= D_{m1} \cap E = \{adf, bcf\}, \\ K_{D_{m2}} &= D_{m2} \cap E = \{acb, bcf\}. \end{aligned}$$

首先考虑（D_1,D_{m1}，D_1^-）。注意，图1中的G_1 与图3中的G_2与（D_1,D_{m1}，D_1^-）是一致的。由控制规范$K_{D_{m1}}$ = {ad f , bc f }，我们有$K_{D_{m1}}$ = {ε, a, b, ad, bc, ad f , bc f }以及$K_{D_{m1}}\Sigma_u$ ={d, ad, bd, add, bcd, ad f d, bc f d}。由于只有ad属于$K_{D_{m1}}$，而$K_{D_{m1}}\Sigma_u$ 中的其他字符串属于D_1^-，条件（9）成立，并且（D_1,D_{m1}，D_1^-）对E是标记信息充分的。事实上，我们可以确认$K_{D_{m1}}$ 相对于一致的被控对象G_1 和G_2 的能控性（若它们可得）。相应地，对每一个与（D_1, D_{m1}，D_1^-）一致的被控对象G，按如下方式构造一个标记非阻塞监督器V_{D_1}：$D_1 \to Pwr(\Sigma_c)$，使得$L_m(V_{D_1}/G) = K_{D_{m1}}$：

$$V_{D_1}(s) = \begin{cases} \{c, e, f\} & \text{if } s = \varepsilon, \\ \{a, b, e, f\} & \text{if } s = b, \\ \{a, b, c, e\} & \text{if } s \in \{ad, bc\}, \\ \{a, b, c, e, f\} & \text{if } s \in \{a, adf, bcf\}, \\ \emptyset & \text{if } s \in \overline{D_1} \setminus \overline{K_{D_{m1}}}. \end{cases}$$

图6展示了所构造的监督器V_{D_1}。

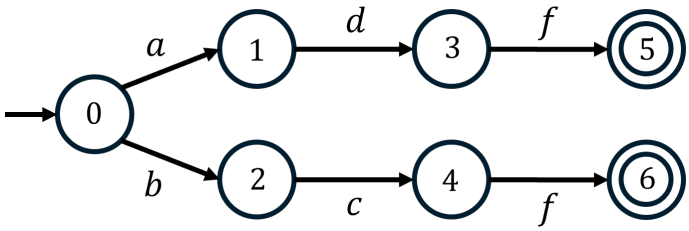


图6. 例3中的监督器V_{D_1}。

接下来，我们考虑 (D_2, D_{m2}, D_2^-)。注意，图1中的 G_1 和图4中的 G_3 均与 (D_2, D_{m2}, D_2^-) 一致。根据控制规范 $K_{D_{m2}} = \{$ acb, bc f $\}$，我们有 $K_{D_{m2}} = \{\varepsilon, a, b, ac, bc, acb, bc f\}$ 和 $K_{D_{m2}}\Sigma_u = \{$d, ad, bd, acd, bcd, acbd, bc f d$\}$。由于 ad, acd $\notin K_{D_{m2}} \cup D_2^-$，$(D_2, D_{m2}, D_2^-)$ 对 E 并非标记信息充分。字符串 acd 属于 $\overline{D_2} \setminus \overline{K_{D_{m2}}}$，因此它不在规范之内，并且每个一致被控对象 G 都能生成该字符串。另一方面，我们没有关于字符串 ad 的信息，因此某些一致被控对象 G 能生成它，而其他被控对象则不能。例如，图1中的 G_1 可以生成 ad，但图4中的 G_3 无法生成它。

为比较考虑和不考虑标记行为的情况，注3引入了不考虑标记行为的数据信息性及其相关定义。

注3：不考虑标记行为时，数据仅由（D，D^-）组成。在这种情况下，基于数据的控制规范定义为

$$K_D := \overline{D} \cap E \quad (E \subseteq \Sigma^*). \tag{14}$$

作为（9）的替代，不考虑标记行为时的信息性判据由下式给出：

$$(\forall s \in \overline{K_D}, \forall \sigma \in \Sigma_u) \quad s\sigma \in \overline{K_D} \cup D^-. \tag{15}$$

监督器 $V'_D : D \to Pwr(\Sigma_c)$ 满足 $L(V_D/G) = K_D$，作为（13）的替代，按如下方式构造：

$$V'_D(s) = \begin{cases} \{\sigma \in \Sigma_c \mid s\sigma \notin \overline{K_D}\} & \text{if } s \in \overline{K_D}, \\ \emptyset & \text{if } s \in \overline{D} \setminus \overline{K_D}. \end{cases} \tag{16}$$

比较考虑和不考虑标记行为的数据信息性判据（即（9）和（15））可知，唯一的区别在于将控制规范 K_{D_m}（基于 D_m）替换为 K_D（基于 D），而结构保持不变。这种相似性源于两种情况下的信息性均通过确保两个条件来确立：如（11）和（12）所示，在观测范围内具备可控性；以及保证不可观测的不可控事件不会发生，这由 D^- 确保。尽管（9）和（15）非常相似，但采用我们的方法构造的监督器（该方法考虑标记行为）与不考虑标记行为时构造的监督器，在所构造的监督器是否保证非阻塞这一点上存在显著差异。以下例4将对此进行说明。

例4：下面使用例3中介绍的相同机器人导航示例，演示不考虑标记行为时的数据信息性和所构造的监督器。假设图1中的受控未知被控对象为 G_1，且给定事件集 $\Sigma = \Sigma_c \cup \Sigma_u$。我们考虑一个有限数据集 (D_1, D_1^-)，它也与例3中的数据集相同。令 $E = \{ae, acb, adf, bcf\}$。则式（14）中的控制规范为

$$K_{D_1} = \overline{D_1} \cap E = \{ae, adf, bcf\}.$$

由于 (15) 在 K_{D_1} 下成立，因此 (D_1, D_1^-) 是信息充分的。相应地，可以构造一个监督器 $V'_{D_1} : D_1 \to Pwr(\Sigma_c)$，使得 $L(V_{D_1}/G) = K_{D_1}$；对于每个与 (D_1, D_1^-) 相一致的被控对象 G，均可按如下方式构造该监督器：

$$V'_{D_1}(s) = \begin{cases} \{c, e, f\} & \text{if } s = \varepsilon, \\ \{a, b, c, f\} & \text{if } s = a, \\ \{a, b, e, f\} & \text{if } s = b, \\ \{a, b, c, e\} & \text{if } s \in \{ad, bc\}, \\ \{a, b, c, e, f\} & \text{if } s \in \{ae, adf, bcf\}, \\ \emptyset & \text{if } s \in \overline{D_1} \setminus \overline{K_{D_1}}. \end{cases}$$

图 7 显示了所构造的监督器 V'_{D_1}。

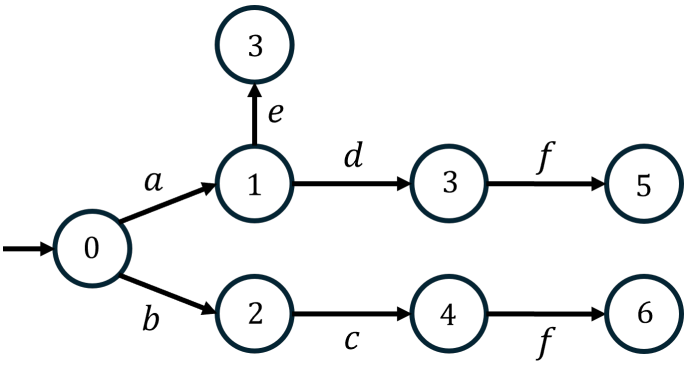


图 7. 例4中的监督器 V'_{D_1}。

比较图 6 和图 7 可以发现，图 7 保留了一条从状态 1 经事件 e 到状态 3 的路径。在机器人导航示例中，这条路径代表一种阻塞情形：智能体未能到达目标。然而，如图 6 所示，通过考虑标记行为，可以消除这类阻塞路径，从而确保只有通向目标的路径才是允许的。这一区别凸显了在数据驱动监督控制中考虑标记行为的重要性，尤其是在必须保证目标导向行为的场景中。另一方面，对于中间行为或部分达成目标已足够、且不要求严格实现目标的任务，不考虑标记行为的监督控制或许更为适用。从这一角度看，考虑标记行为是监督控制中的关键因素，因为它既能确保目标达成行为，又能防止阻塞路径，因此对于实现特定结果至关重要的应用而言，这是一种不可或缺的方法。

C. 标记数据信息性的验证

基于定理1中的充要条件，接下来我们将介绍一种用于检查标记数据信息性的算法。为此，我们首先

定义一种数据驱动自动机。我们用 q_s 表示从自动机的初始状态出发，经字符串s 到达的状态。在数据驱动自动机中，状态与字符串一一对应。

定义 5：（数据驱动自动机）。假设给定事件集 Σ 以及有限数据集 $D,D_m,D^- \subseteq \Sigma^*_$（满足 $D_m \subseteq D$，且 $D \cap D^- = /0$）。则数据驱动自动机定义如下：

$$\hat{G}(\Sigma,D,D_m,D^-) = (\hat{Q},\Sigma,\hat{\delta},\hat{q}_\varepsilon,\hat{Q}_m), \tag{17}$$

其中 $\hat{Q} := \{\hat{q}_s \mid s \in \overline{D \cup D^-}\}$ 为状态集，$\hat{\delta} := \{(\hat{q}_s,\sigma) \to \hat{q}_{s\sigma} \mid s \in D \cup D^-, \sigma \in \Sigma, s\sigma \in D \cup D^-\}$ 为（部分）状态转移函数，$\hat{q}_\varepsilon$为初始状态，$\hat{Q}_m := \{\hat{q}_s \mid s \in D_m\}$ 为标记状态集。此外，给定控制规范 $K_{D_m} = D_m \cap E$（其中 $E \subseteq \Sigma^*$ 为正则语言），我们定义 $Q_K := \{\delta(\hat{q}_\varepsilon,s) \mid s \in K_{D_m}\} \subseteq \hat{Q}$, $Q__ := \{\hat{\delta}(\hat{q}_\varepsilon,s) \mid s \in D^-\} \subseteq \hat{Q}$。

一个数据驱动自动机 $\hat{G}$ 是 D 的一个前缀树自动机：即一个无环自动机，其闭行为为 $L(\hat{G}) = \overline{D \cup D^-}$，标记行为为 $L_m(\hat{G}) = D_m$。根据状态转移函数 $\hat{\delta}$，对于每个字符串 $s \in \overline{D \cup D^-}$，到达的状态 $\hat{q}_s$ 是唯一的。状态子集 Q_K 包含由 K_{D_m} 中字符串到达的那些状态。注意，由于 E 一般不一定是有限语言，为了确定 Q_K，我们首先为 E 构造一个（有限状态）自动机（由于 E 是正则的，这始终可行），然后检查有限集合 D_m 中的每个字符串是否出现在 E 的自动机中。另一方面，状态子集 $Q__$ 包含由 D^- 中字符串到达的那些状态，这意味着向 $Q__$ 的转移表示（未知）被控对象的不可能行为。由于 $D \cap D^- = /0$，我们有 $Q_K \cap Q__ = /0$。需要指出的是，一般而言 $Q_K \cup Q__ \neq \hat{Q}$。此外，$\hat{G}$ 与 (D, D_m, D^-) 不一致，因为 $\hat{G}$ 生成了 D^- 中的字符串：即 $D^- \subseteq L(\hat{G})$。

例5：这里给出例3中与数据集（D_1,D_{m1},D_1^-）和（D_2,D_{m2},D_2^-）相对应的数据驱动自动机 $\hat{G}_1$ （见图8）和 $\hat{G}_2$ （见图9）。为便于清晰展示，图中省略了 $\hat{q}_s$，每个状态内部仅标注下标 s 。图中分别用橙色和蓝色表示状态子集Q_K 和$Q__$。未着色的状态对应于观测数据集D中的字符串，但不属于规范E （因此也不属于K_{D_m}）。

现在，我们已经准备好介绍一种基于数据驱动自动机的标记信息性验证算法。

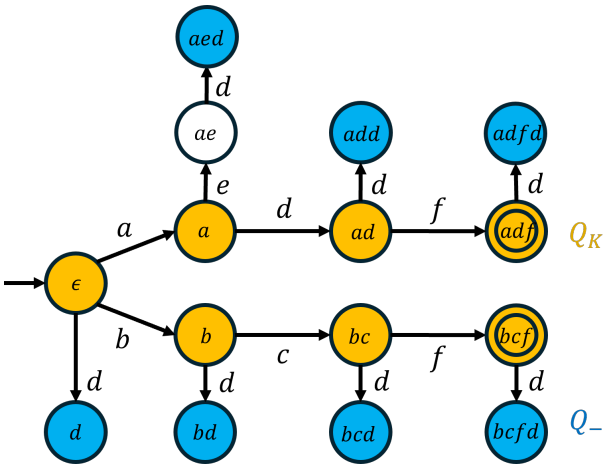


图8. 与（D_1,D_{m1},D_1^-）相对应的数据驱动自动机 $\hat{G}_1$ 。

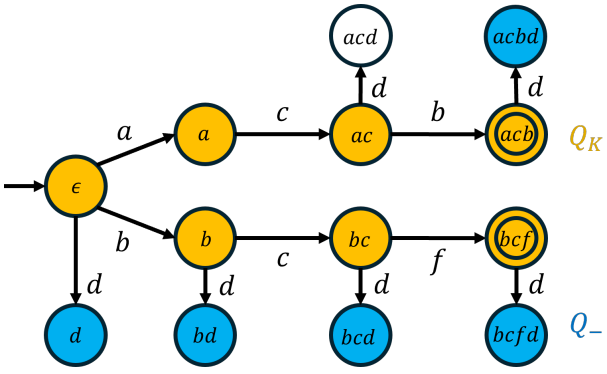


图9. 与（D_2,D_{m2},D_2^-）相对应的数据驱动自动机 $\hat{G}_2$ 。

算法1 检查标记信息性

要求：事件集 $\Sigma = \Sigma_c \cup \Sigma_u$，有限集 $D, D_m, D^-(\subseteq \Sigma^*)$，控制规范 K_D $_m = D_m \cap E$

确保："标记信息充分"或"标记不具信息性"

1: 构造一个数据驱动自动机 $\hat{G}(\Sigma, D, D_m, D^-) = (\hat{Q}, \Sigma, \hat{\delta}, \hat{q}_\varepsilon, \hat{Q}_m)$ 和 Q_K, Q_- (如定义 5 所示)
2: 对于所有 $\hat{q} \in Q_K$，
3: 对所有 $\sigma \in \Sigma_u$，执行
4: if $(\hat{\delta}(\hat{q}, \sigma)!$ 且 $\hat{\delta}(\hat{q}, \sigma) \in \hat{Q} \setminus (Q_K \cup Q_-))$ 或 $(\neg\hat{\delta}(\hat{q}, \sigma)!)$ 则
5: 返回 "标记信息不充分"
6: 中断
7: 结束 if
8: 结束 for
9: 结束 for
10: 返回 "标记信息充分"

在算法1中，通过检查数据驱动自动机中每个状态 Q_K 下的每个不可控事件，来确定 (D, D_m, D^-) 关于 K_{D_m} 的标记数据信息性。如果不可控事件 σ 可以在状态 $\hat{q} \in Q_K$ 下发生，且相应的转移进入 $\hat{Q} \setminus (Q_K \cup Q_-)$，那么该转移包含于 $L(G)$ 中（对于所有与 (D, D_m, D^-) 一致的被控对象 G），但不包含于 K_{D_m} 中，这意味着存在一个 K_{D_m} 中的字符串通过某个不可控事件离开 K_{D_m}。因此，对于每个与 (D, D_m, D^-) 一致的被控对象 G，K_{D_m} 都是不可控的，从而 (D, D_m, D^-) 不是标记信息充分的。如果不可控事件 σ 无法在状态 $\hat{q} \in Q_K$ 下发生，这意味着我们没有关于相应转移的数据或先验知识，因此无法确定该转移是否可由未知真实被控对象 G 生成。因此，(D, D_m, D^-) 不是标记信息充分的。算法1的正确性由下面的命题予以说明。

命题 3：当且仅当 (D, D_m, D^-) 对 E 标记信息充分时，算法1返回"标记信息充分"。

证明：（若）假设 (D, D_m, D^-) 对 E 标记信息充分。由定理1可知，对于每个字符串 $s \in K_{D_m}$ 和每个不可控事件 $\sigma \in \Sigma_u$，式（9）均成立。根据 Q_K 和 Q_- 的定义，$\sigma \in \Sigma_u$ 在所有 $\hat{q}_s \in Q_K$ 处均有定义，且 $\hat{q}_{s\sigma} = \delta(\hat{q}_s, \sigma) \in Q_K \cup Q_-$ 成立。这意味着算法 1 返回 " 标记信息充分 "。

（仅当）假设算法1返回"标记信息充分"。那么，对于所有 $\hat{q}_{s\sigma} = \delta(\hat{q}_s, \sigma) \in Q_K \cup Q_-$，且对于所有 $\sigma \in \Sigma_u$，该条件均成立。若 $\hat{q}_{s\sigma} \in Q_K$，则 $s\sigma \in K_{D_m}$ 成立；若 $\hat{q}_{s\sigma} \in Q^-$，则 $s\sigma \in D^-$ 成立。因此，式（9）成立，且 (D, D_m, D^-) 对 E 标记信息充分。 ■

例6：考虑图8中的数据驱动自动机 $\hat{G}_1$。对于状态 $\hat{q}_a \in Q_K$（橙色）以及（唯一的）$\hat{}$不可控事件 d，满足 $\delta(\hat{q}_a, d) \in Q_K$（橙色）。对于每一个其他状态 $\hat{q}_s \in Q_K$（橙色）以及 $\hat{}$不可控事件 d，满足 $\delta(\hat{q}_s, d) \in Q_-$（蓝色）。因此，算法1返回标记信息充分，这与例3中的结果一致。接下来，考虑图9中的数据驱动自动机 $\hat{G}_2$。对于状态 $\hat{q}_{ac} \in Q_K$（橙色）以及不可控事件 $\hat{} d$，我们看到 $\delta(\hat{q}_{ac}, d)!$ 且 $\delta(\hat{q}_{ac}, d) \notin Q_K \cup Q_-$（因为该转移进入一个白色状态）。因此，算法1返回非标记信息充分，这再次与例3中的结果一致。事实上，基于状态 $\hat{q}_a \in Q_K$ $\hat{}$（橙色）也可以得出相同的结论；此处 $\neg\delta(\hat{q}_a, d)!$，因此算法1返回非标记信息充分。

本节最后讨论数据集 (D, D_m, D^-) 的数量与质量。另一方面，扩大集合 D, D_m（通过增加观测）可以减少与实际对象无法区分的模型数量，也能使更多行为得以实施。另一方面，根据定理1（以及算法1），更大的 D, D_m 意味着需要针对条件 (9) 检查更多字符串，因此更难满足标记数据信息性（除非先验知识数据 D^- 也能相应扩大）。因此，标记数据信息性关注的不只是数据的绝对数量，还关注观测 D, D_m 与先验知识 D^- 之间的匹配质量（正如注2中所指出的，这种质量体现为是否满足 (9)）。如果 (D, D_m, D^-) 对某个规范 E 不具备标记信息充分性，且先验知识 D^- 无法扩大，那么与其考虑为 D, D_m 增加观测，不如寻找一个更小的子集 $K \subseteq K_{D_m}$，使得 (D, D_m, D^-) 可能具备标记信息充分性。下一节将研究这个问题。

IV. 受限标记数据信息性

第IV节讨论数据集（D，D_m，D^-）对于给定规范不具备标记信息充分性的情况，并引入受限标记数据信息性的概念。该概念确保（D，D_m，D^-）对于规范的一个更小子集具备标记信息充分性，从而能够构造一个有效监督器来满足缩减后的规范。

A. K-信息充分性

给定事件集 $\Sigma(= \Sigma_c \cup \Sigma_u)$、有限数据集 D、D_m、$D^- \subseteq \Sigma^*$，以及式（7）中的控制规范 K_{D_m}，若经验证（D、D_m、D^-）对 E 不具有标记信息性，则不存在能够解决问题1的监督器。然而，（D、D_m、D^-）仍有可能对更小的子集 $K \subseteq K_{D_m}$ 具有标记信息性。若情况如此，则可以构造一个有效监督器，对所有与（D、D_m、D^-）相一致的被控对象（包括实际对象）施加较小的规范 K。本节将阐述受限标记数据信息性的概念，其旨在通过将规范限制为更小的子集来建立标记信息

集。受限标记信息性是否成立，取决于所讨论的特定子集K，因此我们将其简称为K-信息充分性，并定义如下。

定义 6：（K-信息充分性）。考虑式 (7) 中的一个特定规范 K^{D_m}，并令 $K \subseteq K^{D_m}$。如果存在一个标记非阻塞监督器 $V_K : D \to Pwr(\Sigma_c)$，使得对于所有与 (D, D_m, D^-) 一致的被控对象 G，均有 $L_m(V_K/G) = K$，则称 (D, D_m, D^-) 是 K-信息充分的；等价地，如果 K 相对于所有与 (D, D_m, D^-) 一致的被控对象 G 都是可控的，也可作此称谓。

根据此定义，如果 (D, D_m, D^-) 对规范 $E \subseteq \Sigma^*$ 已经是标记信息充分的，那么 (D, D_m, D^-) 就是 K^{D_m}-信息充分的（$K^{D_m} = D_m \cap E$）。另一方面，如果 $K = /0$，由于 /0 显然是可控的，(D, D_m, D^-) 总是 /0-信息充分的。然而，强制实现 /0（即空行为）在实践中几乎没有用处，因此下文只考虑非空K。

如果对于给定的K，(D, D_m, D^-)是K-信息充分的，那么可以按照命题 2中的相同方式，为每个与(D, D_m, D^-)一致的被控对象构造出实现K 的标记无阻塞监督器。

对于K-信息充分性的验证，只需对算法1进行直接修改。下面将其作为命题 3的推论予以陈述。

推论 1：假设给定事件集$\Sigma = \Sigma_c \cup \Sigma_u$、有限数据集$D, D_m, D^- \subseteq \Sigma^*$、控制规范 $E \subseteq \Sigma^*$ 以及子集 $K \subseteq K^{D_m}$，其中 K^{D_m}满足式(7)。那么，将Q_K 重新定义为$Q_K := \hat{}\{\delta(\hat{}q_\varepsilon, s) \mid s \in K\}$ 后的算法1，当且仅当(D, D_m, D^-)是K-信息充分的时，返回“标记信息充分”。

我们注意到，一般而言，$(K$-信息充分性 (D, D_m, D^-) 并不意味着对于$K' \subseteq K$ 具有K'-信息充分性（这类似于可控语言的子语言不一定可控这一事实）。例 7 展示了这一点。

例 7：考虑事件集 $\Sigma = \Sigma_c \cup \Sigma_u$，其中 $\Sigma_c = \{a, b\}$ 和 $\Sigma_u = \{d\}$，以及有限数据集 (D_3, D_{m3}, D_3^-)，其中

$$\begin{aligned} D_3 &= \{abd\}, \\ D_{m3} &= \{ab, abd\}, \\ D_3^- &= \{d, ad, abdd\}. \end{aligned}$$

令 $E = \{ab, abd\}$。那么，式（7）中的控制规范为$K^{D_{m3}} = \{ab, abd\}$。相应的数据驱动自动机 $\hat{}G_3$ 如图 10 所示。

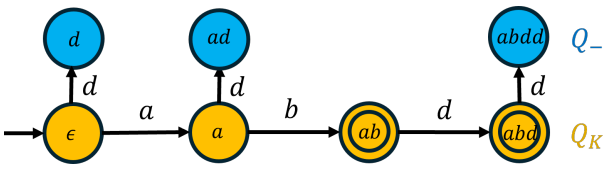


图 10. 与 (D_3, D_{m3}, D_3^-) 对应的数据驱动自动机 $\hat{}G_3$。

通过应用算法1，对于每个状态 $\hat{}q_s \in Q_K$ 以及$\hat{}$ 不可控事件 d，均满足 $\delta(\hat{}q_s, d) \in Q_K \cup Q_-$，因此 (D_3, D_{m3}, D_3^-) 对 E。因此，(D_3, D_{m3}, D_3^-) 是 $K^{D_{m3}}$-信息充分的。然而，令 $K'Dm3 := \{ab\}$(因此 $K'Dm3 \subseteq K^{D_{m3}}$)。按照推论 1 应用修改后的算法 1，返回“标记信息不充分”。这是因为 $\hat{}q_{ab} \in Q_K$，但 $\hat{}\delta(\hat{}q_{ab}, d) \notin Q_K \cup Q_-$。因此，$(D_3, D_{m3}, D_3^-)$ 对 $K'Dm3$ 并非信息充分。

B. 标记可信息化性

在前面的 IV-A 节中，我们针对给定子集$K \subseteq K^{D_m}$定义并检验了 K-信息充分性；本节则探讨是否存在这样的非空子集K。下面将对这一问题进行表述。

问题 2：给定一个事件集$\Sigma = \Sigma_c \cup \Sigma_u$、一个控制规范$E$以及有限数据集$D$，$D_m$，$D^- \subseteq \Sigma^*$，且$K_{D_m} \neq /0$、$D_m \subseteq D$ 并且$D \cap D^- = /0$。判断是否存在一个非空子语言$K \subseteq K^{D_m}$，使得（D，D_m，D^-）是 K-信息充分的。

我们将问题 2 的可解性称为数据（D，D_m，D^-）的一种性质。

定义 7：（标记可信息化性）。对于给定的控制规范E，若存在一个非空子语言$K \subseteq K^{D_m}$，使得（D，D_m，D^-）是 K-信息充分的，则称（D，D_m，D^-）对于该控制规范是标记可信息化的。

在不考虑标记行为的情况下，已有研究提出了可信息化性的判据，并基于该判据给出了验证算法 [17]。然而，考虑标记行为时，无法直接应用类似的判据。

此外，我们还证明了类似判据在标记行为背景下并不适用。具体而言，我们给出了一个注释，引入先前研究中所采用的判据，并通过一个具体示例确认了其局限性。

注释 4：我们引入了在不考虑标记行为 [17]情况下的信息化判据。请注意，以下判据中使用的控制规范是在（14）中引入的（即基于D），而不是在（7）中引入的控制规范。

设事件集 $\Sigma = \Sigma_c \cup \Sigma_u$ 和控制规范E 给定。那么，(D, D^-)对于E 是可信息化的，当且仅当：

$$(\forall s \in \overline{K_D} \cap \Sigma_u^*, \forall \sigma \in \Sigma_u) \quad s\sigma \in \overline{K_D} \cup D^-. \tag{18}$$

例8：考虑事件集 $\Sigma = \Sigma_c \cup \Sigma_u$，其中$\Sigma_c = \{a, b\}$ 和 $\Sigma_u = d$，以及有限数据集(D_4, D_{m4}, D_4^-)，其中

$$\begin{aligned} D_4 &= \{a, db\}, \\ D_{m4} &= \{a, db\}, \\ D_4^- &= \{d, ad, dd\}. \end{aligned}$$

设$E = \{a, db\}$。则（7）中的控制规范为$K_{D_{m4}} = \{a, db\}$。相应的数据驱动自动机$\hat{}G_4$ 如图11所示。

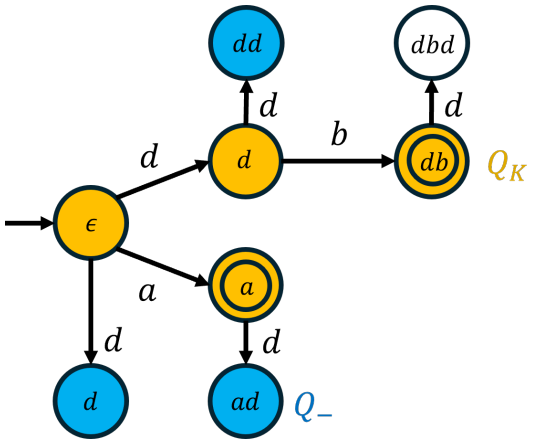


图11. 对应于（D_4,D_{m4}，D_4^-）的数据驱动自动机 $\hat{}G_4$。

首先应用算法1可知，对于状态$\hat{}q_{db} \in Q_K$ 以及不可控事件d，我们看到$\hat{}$ $\hat{}$即 $\delta(\hat{}q_{db}, d)!$且 $\delta(\hat{}q_{db}, d) \notin Q_K \cup Q_-$，因此($D_4, D_{m4}, D_4^-$)对E不是标记信息充分的。接下来，我们验证当$K_D$ 被替换为$K^{D_{m4}}$时，（18）成立。由于$K_{D_{m4}} = \overline{\{\varepsilon, a, d, db\}}$，$K_{D_{m4}} \cap \overline{\Sigma_u^*} = \{\varepsilon, d\}$；并且由于 {d, dd} $\subseteq D^-$，（18）成立。然而，我们得出结论：(D_4, D_{m4}, D_4^-)不是标记可信息化的。这是因为，要禁止状态$\hat{}q_{db}$（它是$\hat{}G_4$，中违反标记信息性的原因），就必须禁止从状态$\hat{}q_d$ 到状态$\hat{}q_{db}$的可控事件上的转移b。因此，如果从状态$\hat{}q_\varepsilon$到状态$\hat{}q_d$ 的转移仍然保留，将导致阻塞，因而该转移也必须被禁止。然而，由于d是不可控事件，它不能被禁止。因此可以得出结论：不存在子语言$K \subseteq K_{D_{m4}}$使得(D_4, D_{m4}, D_4^-)是K-信息充分的。

如例8所示，在所提出的数据驱动标记监督控制中，构造的监督器必须是非阻塞的。因此，与不考虑标记行为的情况相比，给定数据集可标记信息化的条件变得更加严格（具体而言，条件（4.1）虽然必要，但并不充分）。为应对这一挑战，下一章将提出一种算法，通过计算满足（D，D_m，D^-）具有K-信息性的最大子语言$K \subseteq K^{D_m}$，并判断其是否为空，从而确定标记可信息化性。

V. 最少受限标记数据信息性

如果验证数据集（D，D_m，D^-）具有标记可信息化性，则可以确定存在一个非空子集$K \subseteq K^{D_m}$，使得（D，D_m，D^-）是 K-信息充分的。也就是说，（D，D_m，D^-）满足关于K的受限标记信息性。在本节中，我们将进一步探讨如何系统地找到这样的非空K。特别值得关注的是，找到（如果可能）最大的$K_{\sup} \subseteq K^{D_m}$，使得（$D$，$D_m$，$D^-$）关于$K^{\sup}$具有最小受限标记信息性。那么相应的监督器强制实施$K^{\sup}$ 的监督器是最大许可的，因为它允许尽可能大的行为集合。

我们首先定义K^{D_m}的以下子集族；相对于该子集族，数据集（D，D_m，D^-）是受限标记信息性的：

$$I(K_{D_m}) := \{K \subseteq K_{D_m} \mid (D, D_m, D^-) \text{ is K-informative}\}. \tag{19}$$

因此，如果（D，D_m，D^-）是可标记信息化的，则族$I(K^{D_m})$包含一个非空成员。另请注意，由于$K^{D_m}(= D_m \cap E)$是有限的，$I(K^{D_m})$中的成员数也是有限的。

下面的结果至关重要，它断言了如下族命题 4：考虑式 (19) 中的 $I(K^{D_m})$ 族。若 $K_1, K_2 \in I(K^{D_m})$，则 $K_1 \cup K_2 \in I(K^{D_m})$。

证明：根据有限标记信息性的定义，令 $s \in K_1 \cup K_2$ 和 $\sigma \in \Sigma_u$，我们将证明 $s\sigma \in K_1 \cup K_2 \cup D^-$。由于 $s \in K_1 \cup K_2 = K_1 \cup K_2$，因此 $s \in K_1$ 或 $s \in K_2$。考虑前一种情形 $s \in K_1$（后一种情形 $s \in K_2$ 是对称的）。由于 (D, D_m, D^-) 具有 K_1-信息性，我们有

$$s\sigma \in \overline{K_1} \cup D^- \subseteq (\overline{K_1} \cup \overline{K_2}) \cup D^- = \overline{K_1 \cup K_2} \cup D^-.$$

证明完毕。 ■

鉴于命题 4，$I(K^{D_m})$ 族包含唯一的最大成员 $K^{\sup}$，它是该族所有成员的并集：

$$K_{\sup} := \bigcup \{K \mid K \in I(K_{D_m})\}. \tag{20}$$

关于此$K_{\sup} \subseteq K^{D_m}$，数据集（$D$、$D_m$、$D^-$）具有最小受限标记信息性。

对于与（D、D_m、D^-）一致的每个被控对象，都可以按照命题 2中的相同方式构造实施$K^{\sup}$ 的监督器；由于$K^{\sup}$的最大广泛性，该监督器是最大许可的。

既然我们已经证明了最大状态子集$K^{\sup}$的存在性和唯一性，接下来我们将开发一种算法来计算$K^{\sup}$。

A. 非信息状态

为计算$K^{\sup}$，我们首先引入非信息状态这一有用概念。

给定事件集 Σ、控制规范 E以及有限数据集 D、D_m、$D^- \subseteq \Sigma^*$（满足 $D_m \subseteq D$ 且 $D \cap D^- = /0$），按照定义 5 构造相应的数据驱动自动机 $\hat{}G(\Sigma, D、D_m、D^-) = (\hat{}Q\ \Sigma, \hat{}\delta, \hat{}q_\varepsilon, \hat{}Q_m)$，其状态子集 Q_K（对应于 $K_{D_m} = D_m \cap E$）和 Q_-（对应于 D^-）如定义 5所示。我们识别出 Q_K 中违反标记信息性条件（9）的那些状态。这些状态恰好是满足算法1第4行条件的状态。根据该条件，我们给出如下定义。

定义8：（非信息状态）。考虑数据驱动自动机 $\hat{}G(\Sigma, D, D_m, D^-) = (\hat{}Q\ \Sigma, \hat{}\delta, \hat{}q_\varepsilon, \hat{}Q_m)$，其状态子集为 Q_K，

Q_-（如定义 5 所述）。若 $\hat{q} \in Q_K$ 是一个非信息状态，则称之为

$$(\exists \sigma \in \Sigma_u)\ \hat{\delta}(\hat{q},\sigma) \in \hat{Q} \setminus (Q_K \cup Q_-) \quad \text{or} \quad \neg\hat{\delta}(\hat{q},\sigma)! \quad (21)$$

此外，将非信息状态集合定义如下：

$$N(Q_K) := \{\hat{q} \in Q_K \mid \hat{q} \text{ is a non-informative state}\}. \quad (22)$$

式 (22) 中的集合 $N(Q_K)$ 为标记信息性提供了另一种刻画方式（相较于命题 3），如下所述。

命题 5：考虑一个事件集 $\Sigma = \Sigma_c \cup \Sigma_u$、有限数据集 D、$D_m$、$D^- \subseteq \Sigma^*$（满足 $D_m \subseteq$ D、D $\cap D^- = /0$）、一个控制规范 E $\subseteq \Sigma^*$、一个子集 K $\subseteq K_{D_m}$，其中 K_{D_m} 满足式 (7)、相应的数据驱动自动机 $\hat{}$G(Σ, D, D_m,D^-) = ($\hat{}$Q Σ, $\hat{}$δ, $\hat{}$$q_\varepsilon$, $\hat{}$$Q_m$)，以及式 (22) 中的非信息状态集合 N($Q_K$)。则有以下结论：

$$N(Q_K) = \emptyset \text{ if and only if } (D, D_m, D^-) \text{ is marking}$$

对于 E 是信息充分的。

证明：根据算法1，当且仅当第 4 行的条件始终不满足时，$N(Q_K) = /0$；这意味着算法1返回“标记信息充分”。由此，根据命题 3，算法1当且仅当 (D, D_m, D^-) 对 E 是标记信息充分的时返回“标记信息充分”；因此结论成立。 ■

非信息状态集的计算可根据算法1进行调整：新算法不再于识别出第一个非信息状态后立即返回“标记信息不充分”，而是检查 Q_K 中的所有状态是否与 Σ_u中的所有不可控事件相关联，并存储所有识别出的非信息状态。下面的算法2给出了计算$N(Q_K)$的新算法。

算法 2 非信息状态集

要求：事件集 $\Sigma = \Sigma_c \cup \Sigma_u$，有限集$D, D_m, D^-$(⊆ Σ*)，控制规范$K_D$ $_m = D_m \cap E$

确保：非信息状态集$N(Q_K)$

1: 构造 $\hat{}G(\Sigma, D, D_m, D^-) = (\hat{}Q\ \Sigma, \hat{}\delta, \hat{}q_\varepsilon, \hat{}Q_m)$ 和 Q_K, Q_-（如定义 5）
2: $N(Q_K) = /0$
3: 对所有 $\hat{}q \in Q_K$ 执行
4: 对所有 $\sigma \in \Sigma_u$ do
5: 如果 ($\hat{}\delta(\hat{}q\ \sigma)!$ 且 $\hat{}\ \delta(\hat{}q\ \sigma) \in \hat{}Q \setminus (Q_K \cup Q_-)$) 或 $\hat{}(\neg\delta(\hat{}q\ \sigma)!)$ 然后
6: $N(Q_K) = N(Q_K) \cup \{\hat{}q\}$
7: 结束 if
8: 结束 for
9: 结束 for
10: 返回 $N(Q_K)$

算法 2 的正确性可由定义8直接得出。

B. 检查标记可信息化性并计算K^{sup}的算法

在数据驱动自动机 $\hat{}G$ 中引入并识别出非信息状态集合 $N(Q_K)$ 后，根据其定义（定义8），为实现受限标记信息性，必须排除 $N(Q_K)$ 中可达非信息状态的 K_{D_m} 中的任何字符串。要实现最小受限标记信息性，即计算 $K_{\sup} \subseteq K_{D_m}$（如式（20）所示），则必须以尽可能小的干预方式排除（进入 $N(Q_K)$ 的字符串）。

我们提出基于数据驱动自动机 $\hat{}G$ 的结构来计算 (20) 中的 $K^{\sup}$。因此，直观地说，上述排除字符串的做法相当于“避开”数据驱动自动机 $\hat{}G$ 中的状态子集 $N(Q_K)$。为获得最大的$K_{\sup} \subseteq K_{D_m}$，应尽可能靠近 $N(Q_K)$ 来“避开” $N(Q_K)$，并通过移除可控事件实现。这种策略与标准基于模型的监督控制理论中计算最大可控子语言的方法类似。有鉴于此，我们基于经过修改的数据驱动自动机结构构建一个监督控制问题，其规范为避开 $N(Q_K)$，并通过 (??) 中的 supcon函数求解该问题，以获得 (20) 中的 $K^{\sup}$。

具体而言，我们定义以下相关自动机。给定事件集 Σ、控制规范 E 以及有限数据集 D、D_m、$D^- \subseteq \Sigma^*$，按照定义 5 构造相应的数据驱动自动机 $\hat{}$G(Σ, D, D_m,D^-) = ($\hat{}$Q Σ, $\hat{}$δ, $\hat{}$$q_\varepsilon$, $\hat{}$$Q_m$)，其状态子集为 Q_K(对应于 K_{D_m} = $D_m \cap$E) 和 Q_- (对应于 D^-)。首先，通过移除 $Q_{r_1} := \{\hat{}q_s \mid (\forall s \in D \cup D^-)s \notin D\}$，从数据驱动自动机 $\hat{}$G 构造一个子自动机 G_D（即子自动机 $\hat{}$ 的状态集为 $\{\delta(\hat{}q_\varepsilon, s)! \mid \forall s \in D\}$），并移除相应的转移；具体如下

$$G_D = (Q_D, \Sigma, \delta_D, \hat{q}_\varepsilon, \hat{Q}_m) \quad (23)$$

其中 $Q_D := \hat{}Q \setminus Q_{r_1}$ 且 $\delta_D := \hat{}\delta \setminus \{(\hat{}q\ \sigma) \to \hat{}q' \mid \hat{}q \in Q_{r_1}$ 或 $\hat{}q' \in Q_{r_1}\}$。注意，根据定义 5 和 $D_m \subseteq D$ 可知，即使移除了 $\hat{}$Q 中的 Q_{r_1}，标记状态也保持不变。此外，我们有 $L(G_D) = D, L_m(G_D) = D_m$。该自动机 G_D 将作为 supcon函数中的被控对象。

接下来，令 $Q_{r_2} := Q_D \setminus Q_K$，并通过移除$N(Q_K) \cup Q_{r_2}$及相应转移，从$G_D$ 构造子自动机S_D；即

$$S_D = (Q_S, \Sigma, \delta_S, \hat{q}_\varepsilon, Q_{S,m}) \quad (24)$$

其中 $Q_S = Q_D \setminus (N(Q_K) \cup Q_{r_2}) = Q_K \setminus N(Q_K)$ 且 $\delta_S := \delta_D \setminus \{(\hat{}q\ \sigma) \to \hat{}q' \mid \hat{}q \in N(Q_K) \cup Q_{r_2}$ 或 $\hat{}q' \in N(Q_K) \cup Q_{r_2}\}$, $Q_{S,m} := \hat{}Q_m \setminus N(Q_K) \cup Q_{r_2}$。此自动机 S_D 将作为 supcon函数中的规范。注意，由于 $Q_S = Q_K \setminus N(Q_K)$，$q \in Q_S$ 中的每个状态都满足否定

(21) 的:

$$(\forall \sigma \in \Sigma_u)\ \hat{\delta}(q,\sigma)! \text{ and } \hat{\delta}(q,\sigma) \in Q_K \cup Q_{-}. \qquad (25)$$

令 (23) 中的 G_D 为被控对象，(24) 中的S_D 为规范。使用 supcon函数计算得到最大允许监督器，它也是一个自动机（实际上，在本例中是S_D的子自动机）：

$$P_D = \text{supcon}(G_D, S_D) = (Q_P, \Sigma, \delta_P, \hat{q}_\varepsilon, Q_{P,m}) \qquad (26)$$

其中 $Q_P \subseteq Q_S$, $\delta_P \subseteq \delta_S$ 且 $Q_{P,m} \subseteq Q_{S,m}$ 满足

$$L_m(P_D) = \sup C(L_m(S_D))$$

其中

$$C(L_m(S_D)) = \{K \subseteq L_m(S_D) \mid \overline{K}\Sigma_u \cap L(G_D) \subseteq \overline{K}\}. \qquad (27)$$

.

我们将上述过程总结为以下算法。所得的 $L_m(P_D)$ 正是我们要找的（20）中的最大 K^{sup} 。此外，若 $L_m(P_D) = K_{sup} \neq /0$，则 (D, D_m, D^-) 是 K^{sup}-信息的，因此 (D, D_m,D^-) 是标记可信息化的。另一方面，若 $L_m(P_D) = K_{sup} = /0$，则不存在非空子集使得 (D, D_m, D^-) 具有受限标记信息性，因此 (D, D_m, D^-) 也不可标记可信息化。因此，以下算法3的非空性可用于检验标记可信息化性。

算法3 检验标记可信息化性并计算 K^{sup}要求：事件集 $\Sigma = \Sigma_c \cup \Sigma_u$、有限集 D、$D_m$、$D^-(\subseteq \Sigma^*)$、控制规范 $K_{D_m} = D_m \cap E$ 确保：“标记可信息化”且 $L_m(P_D)$，或“不可标记可信息化” 1：构造数据驱动自动机 $\hat{}G(\Sigma, D, D_m, D^-) = (\hat{}Q\ \Sigma, \hat{}\delta, \hat{}q_\varepsilon, \hat{}Q_m)$ 以及 Q_K、Q_-（如定义 5 所示） 2：按（23）构造 $\hat{}G$ 的子自动机 G_D 3：通过算法2计算非信息状态集 $N(Q_K)$ 4：按（24）构造 G_D 的子自动机 S_D 5：按（26）计算 $P_D = \text{supcon}(G_D, S_D)$ 6：若 $L_m(P_D) \neq /0$，则 7：返回“标记可信息化”以及 $L_m(P_D)$ 8：否则 9：返回“不可标记可信息化” 10：结束若

命题 6：以下陈述对算法3成立：1. 算法3返回的语言$L_m(P_D)$满足$L_m(P_D) = K^{\sup}$ 于(20)中。2. 算法3返回“标记可信息化”，当且仅当(D, D_m, D^-)对E是“标记可信息化”的。

证明：证明分两部分进行。在第1部分中，我们确立语言$L_m(P_D)$由算法3返回并满足$L_m(P_D) = K^{\sup}$ （见(20)）。我们提出如下关键断言：

$$C(L_m(S_D)) = I(K_{D_m}).$$

即，(27)中$L_m(S_D)$的所有可控子语言构成的族$C(L_m(S_D))$，与(19)中K^{D_m}的所有子集构成的族$I(K^{D_m})$完全相同，对这些子集而言(D, D_m, D^-)是受限标记信息性的。在此论断下，我们推导出

$$L_m(P_D) = \sup C(L_m(S_D)) = \bigcup\{K \mid K \in C(L_m(S_D))\} = \bigcup\{K \mid K \in I(K_{D_m})\}$$

最后一个等式来自(20)，从而我们的结论得以确立。

现在我们证明该论断。首先，设$K \in C(L_m(S_D))$，这意味着$K_{D_m} \subseteq L(S_D)$且$K\Sigma_u \cap L(G_D) \subseteq \overline{K}$。由(24)中$S_D$的定义，$Q_S \subseteq Q_K$，从而$L_m(S_D) \subseteq K_{D_m}$，这进而蕴含$K \subseteq K_{D_m}$。为了证明$K \in I(K_{D_m})$，我们必须证明(D, D_m,D^-)是K-信息充分的。为此，设$s \in K$且 $\sigma \in \Sigma_u$。若$s\sigma \in L(G_D)$，则由$K\Sigma_u \cap L(G_D) \subseteq K$可得$s\sigma \in K$。另一方面，若$s\sigma \notin L(G_D)$，我们推出$s\sigma \in D^-$。为看出这一点，考虑状态$q_s$，它属于$Q_S = Q_K \setminus N(Q_K)$。这意味着(25)对$q_s$成立。但$\delta(q_s, \sigma)$不可能属于$Q_K$，因为$s \notin L(G_D) = D \subseteq K_{D_m}$。因此只可能 $\delta(q_s, \sigma) \in Q_-$，这蕴含$s\sigma \in D^-$。综上所述，对任意$s \in K$以及任意 $\sigma \in \Sigma_u$，我们有$s\sigma \in K \cup D^-$，这意味着(D, D_m, D^-)是K-信息充分的。因此，$K \in I(K_{D_m})$，且 $C(L_m(S_D)) \subseteq I(K_{D_m})$。

下面仍需证明 $I(K_{D_m}) \subseteq C(L_m(S_D))$。令 $K \in I(K_{D_m})$，这意味着 $K \subseteq K_{D_m}$，且 (D, D_m,D^-) 具有 K-信息充分性。考虑任意字符串 $s \in K \subseteq K_{D_m}$。由于 (D, D_m, D^-) 具有 K-信息充分性，状态 $q_s \in Q_K \setminus N(Q_K) = Q_S$。因此 $s \in L_m(S_D)$，进而有 $K \subseteq L_m(S_D)$。为证明 $K \in C(L_m(S_D))$，我们需要证明 K 关于 G_D 是可控的。为此，令 $s \in K$ $\sigma \in \Sigma_u$，并假设 $s\sigma \in \overline{L(G_D)} = \overline{D}$。再次由 (D, D_m,D^-) 的 K-信息充分性可知，$s\sigma \in \overline{K} \cup D^-$。但我们已假设 $s\sigma \in \overline{D}$，因此不可能有 $s\sigma \in D^-$（根据 $\overline{D} \cap D^- = /0$）。所以 $s\sigma \notin K$，这意味着 K 关于 G_D 是可控的。因此 $K \in C(L_m(S_D))$，且 $I(K_D) \subseteq C(L(S_D))$。

鉴于上述内容，我们已经证明了我们的主张：$I(K^{D_m}) = C(L_m(S_D))$，第 1 部分的证明至此完成。

接下来，我们进入第2部分，证明算法3返回“标记可信息化”当且仅当 (D, D_m, D^-) 对于 E 是“标记可信息化”。根据命题6的第一条陈述，算法3正确计算出 K_{D_m} 的最大子集，使得数据集 (D, D_m, D^-) 具有最小受限标记信息性。如果算法3返回“标记可信息化”，那么 $L_m(P_D) = K_{sup} \neq /0$，这进而意味着 (D, D_m,D^-) 是 K^{sup}-信息充分的，因此 (D, D_m, D^-) 是标记可信息化的。另一方面，如果算法3返回“标记不可信息化”，那么 $L_m(P_D) = K^{sup} = /0$，这进而意味着不存在空子集使得

(D, D_m, D^-) 具有受限标记信息性，这意味着 (D, D_m, D^-) 不可标记可信息化。第2部分的证明至此完成。 ■

我们为算法3提供一个说明性示例。

示例9：考虑事件集 $\Sigma = \Sigma_c \cup \Sigma_u$，其中$\Sigma_c = \{a、b、c、e、f\}$和 $\Sigma_u = \{d\}$，以及有限日期集 $(D_5, D_{m5}、D_5^-)$，其中

$$\begin{aligned} D_5 &= \{abe, acfd, aed\}, \\ D_{m5} &= \{ab, abe, acf, aed\}, \\ D_5^- &= \{d, ad, acd, aedd\}. \end{aligned}$$

令 $E = \{ab、acf、aed\}$。则式（7）中的控制规范为 $K_{D_{m4}} = \{ab、acf、aed\}$。相应的数据驱动自动机 $\hat{}G_5$如图12所示。

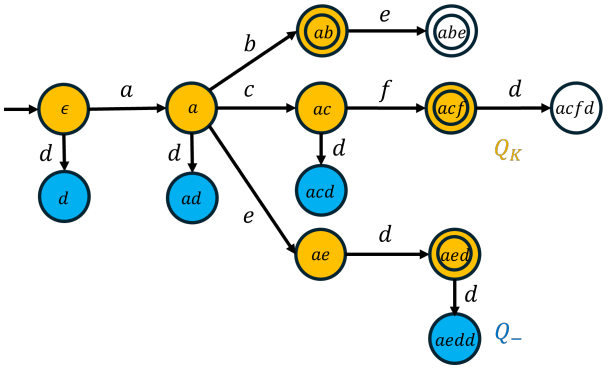


图12. 对应于 $(D_5, D_{m5}、D_5^-)$ 的数据驱动自动机 $\hat{}G_5$ 。

应用算法1可确定，$(D_5, D_{m5}、D_5^-)$ 对于 E“标记信息不充分”。然而，$(D_5, D_{m5}、D_5^-)$ 可能“标记可信息化”，并且可能存在 $K^{D_{m5}}$ 的非空子语言，使得 $(D_5, D_{m5}、D_5^-)$ 具有有限标记信息的性质。接下来，我们应用算法3来检查标记可信息化性，并计算出最大的此类子语言。

构造数据驱动自动机 $\hat{}G_5$ 的算法3第1行已完成，如图12所示。第2行通过从 $\hat{}G_5$ 中移除$Q_{r1} = Q_-$ 中的状态及相关转移（如式（23）所示）来计算子自动机G^{D_5}；结果如图13所示。

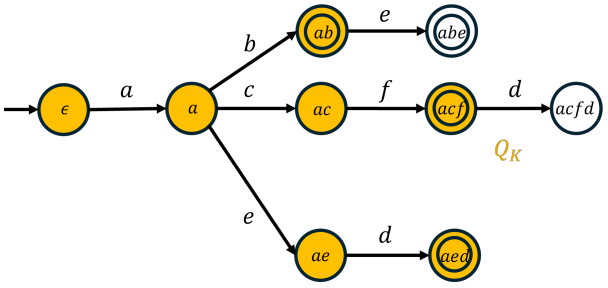


图13. 算法3第2行得到的子自动机G^{D_5} 。

第 3 行应用算法 2 推导非信息状态集合：$N(Q_K) = \{\hat{}q_{ab}, \hat{}q_{acf}\}$。原因是这两个状态属于非信息状态，原因如下：对于 $\hat{}q_{ab} \in Q_K$ 和不可控事件 d，$\neg\delta(\hat{}q_{ab}, d)!$；对于 $\hat{}q_{acf} \in Q_K$ 和不可控事件 d，$\delta(\hat{}q^{acf}, d)!\ \hat{}$且 $\delta(\hat{}q^{acf}, d) \notin Q_K \cup Q_-$。第 4 行通过从 G^{D_5} 中移除 <style id='N'>N(Q_K) 中的两个状态和 Q_{r2} 中的两个状态（以白色标示），并移除相关转移，构造子自动机 S^{D_5} ；结果见图 14。

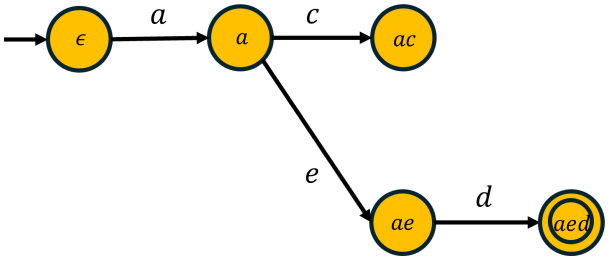


图14. 算法3第4行得到的子自动机S^{D_5} 。 m 3.

最后，第5行计算最大许可监督器$P^{D_5} = \text{supcon}(G^{D_5}, S^{D_5})$；结果如图15所示。

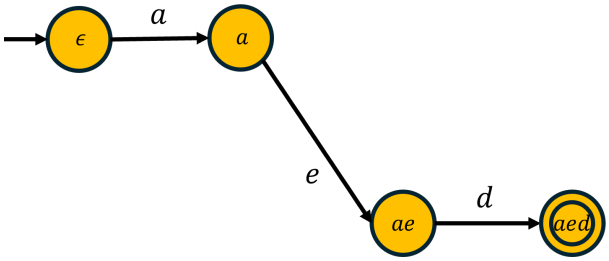


图15. 算法3第5行得到的监督器P^{D_5} 。

与S^{D_5}相比，$\hat{}q_{ac}$ 被移除。这是因为要确保监督器P^{D_5} 是非阻塞的（$\hat{}q_{ac}$ 是可达的，但不是共可达的）。因此，最终结果是（D_5, D_{m5}, D_5^-）具有标记可信息化性质，并且L_m（P^{D_5}）$= \{aed\}$ 是$K^{D_{m5}}$的最大子集，使得数据集（D_5, D_{m5}, D_5^-）具有最小受限标记信息性。

基于$K^{\text{sup}} = L_m(P^{D_5})$，我们可以构造相应的监督器$V_{\text{sup}} : D_5 \to Pwr(\Sigma_c)$，使得$L_m(V_{\text{sup}}/G) = K^{\text{sup}}$ 对于每一个被控对象G ，其与(D_5, D_{m5}, D_5^-)一致，如下：

$$V_{\text{sup}}(s) = \begin{cases} \{b, c, e, f\} & \text{if } s = \varepsilon, \\ \{a, b, c, f\} & \text{if } s = a, \\ \{a, b, c, e, f\} & \text{if } s \in \{ae, aed\}, \\ \emptyset & \text{if } s \in \overline{D_5} \setminus \overline{K_{\text{sup}}}. \end{cases}$$

VI. Conclusions

在本文中，我们研究了离散事件系统的标记数据驱动监督控制。我们提出了标记数据信息性、标记可信息化性和最小受限标记信息性等新概念，并基于一种新颖的数据驱动自动机结构，开发了相应的验证和综合算法

。下面总结了如何使用这些概念和算法：给定一个数据集（D，D_m，D^-）和一个规范E，首先应用算法1验证（D，D_m，D^-）对于E是否具有信息性。如果是，我们便可以构造一个监督器来实现$K^{D_m} = D_m \cap E$。如果否，则应用算法3验证（D，D_m，D^-）对于E是否可信息化。如果是，我们可以得到K^{D_m}的最大子集，使得（D，D_m，D^-）具有最小受限信息性。如果算法3返回否，则不存在能够实现K^{D_m}任何子集的监督器。在这种情况下，可以考虑使用算法??和算法??来收集更多数据，使其相对于更大的语言具有最小受限标记信息性。

为验证所提出的用于标记监督控制的数据驱动方法，我们在典型示例以及涉及“未知环境探索”的场景上进行了实验。结果表明，被控对象的复杂度以及不可控事件的类型会影响实现标记信息性所需的D^-质量。随着被控对象变得更加复杂，或不可控事件数量增加，需要更高质量的D^-，才能使数据集达到标记信息充分。

未来工作的一个有前景的方向是探索如何进一步改进和扩展所提出的数据驱动方法。这包括研究能否在利用已构建的数据驱动自动机的同时，使用新获得的数据有效构建监督器。此外，探索放宽当前假设——例如要求考虑所有不可控事件，除非它们通向D^-——并通过强制来抢先处理不可控事件，仍是一个开放性问题 [15]。进一步地，将信息性从基本可控性扩展到其他关键性质，如可观测性、可诊断性和不透明性，是后续研究的一个重要方向 [20]。

参考文献

[1] S. Takai 和 T. Suzuki，“离散事件系统”，《测量与控制》，第46卷，第4期，第248–254页，2007年。[2] C. G. Cassandras 和 S. Lafortune，《离散事件系统导论》。施普林格，2008年。[3] K. Cai 和 W. Wonham，“离散事件系统的监督控制”，载于《系统与控制百科全书》。施普林格，2021年，第2245–2253页。

[4] P. Ramadge 和 W. Wonham，“一类离散事件过程的监督控制”，《SIAM控制与优化杂志》，第25卷，第1期，第206–230页，1987年。[5] W. Wonham 和 K. Cai，《离散事件系统的监督控制》。施普林格，2019年。[6] W. Wonham 和 P. Ramadge，“给定语言的最大可控子语言”，《SIAM控制与优化杂志》，第25卷，第3期，第637–659页，1987年。[7] W. Wonham、K. Cai 和 K. Rudie，“离散事件系统的监督控制：简史”，《Annual Reviews in Control》，第45卷，第250–256页，2018年。[8] S. Takai，“离散事件系统的监督控制”，《IEICE Fundamentals Review》，第7卷，第4期，第317–325页，2014年。[9] A. Karimi、O. Ghorbani、R. Tashakkori、S. H. R. Pasandideh 和 M. Jasemi，“确定冷藏仓库中食品的最优订货批量、再订货点和质量特征：数据驱动优化方法”，arXiv预印本，第arXiv–2401页，2024年。[10] M. Schrum、E. Sumner、M. Gombolay 和 A. Best，“Maveric：个性化自动驾驶的数据驱动方法”，《IEEE机器人学汇刊》，2024年。[11] K. Yasukochi，“技术的指数级发展与人类增强：加速与整合之外是什么”，《Bulletin of the Institute of Liberal Arts》，九州工业大学，第7期，第1–18页，2023年。[12] L. Chen、C. Xia、Z. Zhao、H. Fu 和 Y. Chen，“人工智能驱动的传感技术”，《Sensors》，第24卷，第10期，第2958页，2024年。[13] M. Konishi、T. Sasaki 和 K. Cai，“通过深度强化学习和监督控制实现高效安全控制——多机器人仓库自动化案例研究”，《IFAC-PapersOnLine》，第55卷，第28期，第16–21页，2022年。[14] T. Steentjes、M. Lazar 和 P. Van den Hof，“关于数据驱动控制：具有交叉协方差界的含噪输入输出数据的信息性”，《IEEE控制系统快报》，第6卷，第2192–2197页，2021年。[15] C. Gu、C. Gao 和 K. Cai，“具有强制事件的离散事件系统的数据驱动监督控制”，《IFAC-PapersOnLine》，第58卷，第1期，第120–125页，2024年。[16] K. Cai，“离散事件系统的数据驱动监督控制”，《系统、控制与信息工程师学会汇刊》（ISCIE），第66卷，第9期，第359–364页，2022年。[17] T. Ohtsuka、K. Cai 和 K. Kashima，“离散事件系统数据驱动监督控制的数据信息性”，载于第62届IEEE决策与控制会议（CDC），2023年，第6923–6928页。[18] K. Fuchiwaki 和 K. Cai，“离散事件系统的标记数据信息性与数据驱动监督控制”，载于《第67届自动控制联合会议论文集》。自动控制联合会议，2024年，第102–107页。[19] J. Hopcroft、R. Motwani 和 J. Ullman，《自动机理论、形式语言与计算导论》。Pearson/Addison Wesley，2007年。[20] C. N. Hadjicostis，《离散事件系统中的估计与推断》。施普林格，2020年。