

态势估计中一种基于最大后验概率估计的时间推理方法^{*}

姚春燕 郁文贤 庄钊文

(国防科技大学自动目标识别重点实验室 长沙 410073)

摘 要 态势估计中统计时间推理在许多应用中非常重要。Kirillov 的基于极大似然估计 (Maximum Likelihood Estimation, MLE) 的推理方法将未知时间变量看作常数, 忽略了它的先验信息, 估计方差较大。针对这一问题, 本文首先建立了已知时间信息和未知时间变量之间的关系模型, 这一模型可用来解释 Kirillov 的方法; 然后在这一模型下, 将未知时间变量扩展为随机变量, 采用基于最大后验概率估计 (Maximum a Posteriori Estimation, MAP) 的方法进行统计时间推理。对两种推理算法的性能进行了分析和比较, 发现在较宽的范围内, 基于 MAP 的方法性能优于基于 MLE 的方法。

关键词 统计时间推理, 态势估计, 极大似然估计, 最大后验概率估计

分类号 O212

A Temporal Reasoning Method Based on the Maximum Posteriori Estimation in Situation Assessment

Yao Chunyan Yu Wenxian Zhuang Zhaowen

(ATR Laboratory, NUDT, Changsha, 410073)

Abstract Stochastic temporal reasoning in situation assessment (SA) is very important in many applications. The approach based on Maximun Likelihood Estimation (MLE) treats the unknown temporal variable as a constant, which doesn't use a priori information and generates a larger estimate variance. In this paper, the relation model of known temporal information and unknown temporal variable has been established, which can also be used for MLE-based method. In the model, the reasoning algorithm about time instants has been derived from treating the unknown temporal variable as random variable and introducing MAP estimation into temporal reasoning. The performance analysis between MAP-based method and MLE-based method shows that under some conditions, the estimate variance of MAP-based method is lower than that of the MLE-based method, and we have given these conditions in experiments.

Key words stochastic temporal reasoning, situation assessment, MLE-based method, MAP-based method

态势估计 (Situation Assessment, SA) 作为军事数据融合系统中的二层处理, 根据不断到来的数据逐步达到对敌方意图的辨别, 可将它归为一个多假设动态分类问题。态势假设可分解成由各子任务 (或称为事件) 按一定逻辑关系组成的层次结构, 其中一些子任务的输出可以作为其他子任务的输入。子任务相同而其间的时间关系不同的态势, 不是同一个态势。因此很有必要将时间推理引入态势估计。SA 中时间推理的功能主要有两个, 一个是根据已知时间变量和时间关系对未知时间变量进行不确定估值; 另一个是确定观察到的时间关系与已知的时间关系是否匹配。本文针对时间变量为时间点的情形, 讨论第一个功能。由于事件发生时间的固有不确定性、以及专家知识含有的不确定性、观测误差的存在等因素的影响, 使得将不确定性时间推理引入态势估计很有必要。

^{*} 1998 年 4 月 8 日收稿
第一作者: 姚春燕, 女, 1972 年生, 博士生

Allen 提出了基于时间区间的十三种时间关系^[1], Dubois 和 Prade 提出了应用可能性理论处理时间知识的理论框架^[2], Kirillov 提出了用于态势估计的统计时间推理方法, 我们称之为基于 MLE 的方法^[3]。这种方法将未知时间变量看作一个未知常数。实际上, 态势发展过程中, 在不考虑观测误差的情况下, 许多事件发生的时间本身具有一定随机性。为解决这一问题, 需要将未知时间变量扩展为一个随机变量。

本文的第 1 部分建立了已知时间信息和未知时间变量之间的关系模型, 并以前向推理为例, 给出了基于 MAP 的时间推理方法, 然后说明了这种方法在后向推理中的应用; 第 2 部分对两种推理算法的性能进行了分析和比较; 最后一部分是结论, 并指出将来的研究方向。

1 基于 MAP 的时间推理方法

我们以谓词 $T(t, t_0)$ 表示一个时间关系, $t_k, k = 0, 1, 2, \dots, n$ 表示时间变量, $k > 0$ 时, t_k 表示子任务各输入事件发生的时间, $k = 0$ 时, t_k 表示子任务输出事件发生的时间, $t = (t_1, \dots, t_n)$ 是时间变量的有序集, 所有的时间变量均为不确定的。我们将时间轴上任一点 T_k , 以有序实数对 $\hat{T}_k = (\hat{\tau}_k, \sigma_{\tau k}^2)$ 表示, 其中 $\hat{\tau}_k$ 是时间点 T_k 的估值, $\sigma_{\tau k}^2$ 是估值的方差, $\hat{T}_k$ 表示对 T_k 的不确定估值。Kirillov 建立了一个统计时间关系——“预测”: $F(t_k, \Delta t_{kj}, \sigma_{\tau k}^2, t_j)$ 。并将它的含义解释为: 对于给定的时间点 T_k , 可以计算另一个时间点 T_{jk} 的不确定估计值 $\hat{T}_{jk} = (\hat{\tau}_{jk}, \sigma_{\tau jk}^2)$, 其中,

$$\hat{\tau}_{jk} = \hat{\tau}_k + \Delta t_{kj} \quad (1)$$

$$\sigma_{\tau jk}^2 = \sigma_{\tau k}^2 + \sigma_{\tau kj}^2 \quad (2)$$

这里假设专家提供的时间点估值 $\hat{\tau}_k$ 和时间距离值 Δt_{kj} 是统计独立的; 并且当 $k \neq i$ 和 $j \neq l$ 时, 任意两个随机变量 Δt_{kj} 和 Δt_{il} 是统计独立的; 所有的随机变量都服从高斯分布。在以上假设下, 预测关系 F 具有传递性 ((3) 式) 和对称性 ((4) 式):

$$\forall (i, j, k), F(t_i, \Delta t_{ij}, \sigma_{\tau ij}^2, t_j) \quad F(t_j, \Delta t_{jk}, \sigma_{\tau jk}^2, t_k) \rightarrow F(t_i, \Delta t_{ij} + \Delta t_{jk}, \sigma_{\tau ij}^2 + \sigma_{\tau jk}^2, t_k) \quad (3)$$

$$\forall (j, k), F(t_j, \Delta t_{jk}, \sigma_{\tau jk}^2, t_k) \rightarrow F(t_k, -\Delta t_{jk}, \sigma_{\tau jk}^2, t_j) \quad (4)$$

由于“预测”关系 F 表示可获得 t_j 的不确定估值的途径之一, 因此, 可以将 F 称为关于 t_j 的一个部分预测关系; 而将 T_{jk} 称为关于 t_j 的一个部分预测值。将预测关系 F 作为最基本的时间关系的专家知识模型, 可将 $\forall T(t, t_0)$ 分解为由带有不同参数的预测关系 F 相“与”而组成的形式。在前向推理中, $T(t, t_0)$ 可分解为 (5) 式。

$$T(t, t_0) \equiv \bigwedge_{k=1}^n F(t_k, \Delta t_{k0}, \sigma_{\tau k0}^2, t_0) \quad (5)$$

我们将集合 $\{1, 2, \dots, n\}$ 记为 M , (5) 式表示, $t^0, t^1, t^2, \dots, t^n$ 之间的时间关系由 n 个关于 t^0 的预测关系组成。假设 $\forall k \in M$, 给定 t_k 的一个不确定估计值 $\hat{T}_k = (\hat{\tau}_k, \sigma_{\tau k}^2)$ 和 Δt_{k0} 的一个不确定估计值 $(\Delta t_{k0}, \sigma_{\tau k0}^2)$, 那么可获得 t_0 的 n 个相互独立的部分预测值, 由此进一步得到 t_0 的不确定估计值。下面, 我们首先建立不确定时间点的时间推理模型, 然后, 运用基于 MAP 的方法, 推导出 t_0 的不确定估计值。

1.1 已知时间信息与未知时间变量的关系模型

由 (5) 式, 可获得 t_0 的 n 个相互独立的部分预测值。将这 n 个部分预测值记为: $(\hat{\tau}_{0k}, \sigma_{\tau 0k}^2)$, $\forall k \in M$, 由 (1) 式、(2) 式可以得到:

$$\hat{\tau}_{0k} = \hat{\tau}_k + \Delta t_{k0} \quad (6)$$

$$\sigma_{\tau 0k}^2 = \sigma_{\tau k}^2 + \sigma_{\tau k0}^2 \quad (7)$$

观测不确定性的影响, 使得 $\hat{\tau}_k$ 与 t_k 的真实值有一定偏差 $n_{\tau k}$, Δt_{k0} 与 Δt_{k0} 的真实值有一定偏差 $n_{\tau k0}$, 即 $\forall k \in M$, 有:

$$\hat{\tau}_k = t_k + n_{\tau k} \quad (8)$$

$$\Delta t_{k0} = \Delta t_{k0} + n_{\tau k0} \quad (9)$$

将 (8) 式、(9) 式代入 (6) 式, 得到: $\hat{\tau}_{0k} = (t_k + n_{\tau k}) + (\Delta t_{k0} + n_{\tau k0}) = t_0 + n_{\tau k} + n_{\tau k0}$

令 $n_{0k} = n_{\tau k} + n_{\tau k0}$, 则

$$\hat{\tau}_{0k} = t_0 + n_{0k} \quad (10)$$

记 $\hat{\tau} = (\hat{\tau}_{01}, \dots, \hat{\tau}_{0n})$, $\mathbf{H} = [1, \dots, 1]^T_{1 \times n}$, $\mathbf{n} = (n_{01}, \dots, n_{0n})$, 则

$$\hat{\tau} = \mathbf{H} \cdot t_0 + \mathbf{n} \quad (11)$$

(11) 式即为已知的部分预测值 $\hat{\tau}$ 与未知时间变量 t_0 的关系模型。在这一模型中, 若将 t_0 看作一个常数, 并假设 $\mathbf{n}$ 是均值为零, 协方差矩阵为 $\text{Cov}(\mathbf{n}) = \text{diag}(\sigma_{n01}^2, \sigma_{n02}^2, \dots, \sigma_{n0n}^2)$, 服从高斯分布的随机矢量, 然后根据 (12) 式计算出 t_0 的 MLE 值, 即为 Kirillov 的计算结果。由于实际情形中, t_0 具有一定随机性, 把它看作常数会产生较大的估计方差, 因此我们把 t_0 看作一个高斯分布的随机变量; 假设 $\mathbf{n}$ 是一个均值为零、协方差矩阵为 $\text{Cov}(\mathbf{n}) = \text{diag}(\sigma_{n01}^2, \sigma_{n02}^2, \dots, \sigma_{n0n}^2)$ 、服从高斯分布的随机矢量, 且 $\forall k \in M, n_{0k}$ 与 t_0 相互独立。此外, 我们进一步假设 $\exists j \in M, \sigma_{n0j}^2$ 已知。这一个假设条件在许多应用背景下可以得到满足^[3]。下面, 我们采用基于 MAP 的估计方法对 t_0 进行不确定估值。

1.2 基于 MAP 的时间推理方法

我们首先估计 t_0 的先验分布密度 $P(t_0)$, 然后计算 $\hat{\tau}$ 的条件分布密度 $P(\hat{\tau} | t_0)$ 。

由于 $\forall k \in M, E[n_{0k}] = 0$, 所以, $\forall k, E[t_0] = E[\hat{\tau}_{0k}] \triangleq \theta_0$ 。则可求得似然函数:

$$L(\hat{\tau}, \theta) = \frac{1}{(2\pi)^{n/2} (\sum_{k=1}^n \sigma_{\tau_{0k}}^2)^{1/2}} \exp \left[- \sum_{k=1}^n \frac{(\hat{\tau}_{0k} - \theta)^2}{2\sigma_{\tau_{0k}}^2} \right] \quad (12)$$

$$\text{令 } \frac{\partial L}{\partial \theta} = 0, \text{ 则} \quad \hat{\theta} = \frac{\sum_{k=1}^n \hat{\tau}_{0k} \sigma_{\tau_{0k}}^{-2}}{\sum_{k=1}^n \sigma_{\tau_{0k}}^{-2}} = E[t_0] \quad (13)$$

$$\text{进一步可以得到:} \quad \sigma_{\hat{\theta}}^2 = \frac{1}{\sum_{k=1}^n \sigma_{\tau_{0k}}^{-2}} \quad (14)$$

由 A 中的假设可得到: $\sigma_{\hat{\theta}}^2 = \sigma_{\tau_{0j}}^2 - \sigma_{n0j}^2$ 。且:

$$\forall k \in M, \sigma_{\tau_{0k}}^2 = \sigma_{t_0}^2 + \sigma_{n0k}^2$$

$$\text{所以} \quad \forall k \neq j, \sigma_{n0k}^2 = \sigma_{\tau_{0k}}^2 - \sigma_{t_0}^2 \quad (15)$$

这样, 可以得到 t_0 的先验分布为:

$$P(t_0) = \frac{1}{2\pi\sigma_{t_0}^2} \exp \left[- \frac{(t_0 - \hat{\theta})^2}{2\sigma_{t_0}^2} \right] \quad (16)$$

根据随机变量的独立性假设和高斯分布假设, 由 (11) 式, 可得到 $\hat{\tau}$ 的条件概率分布:

$$P(\hat{\tau} | t_0) = \frac{1}{(2\pi)^{n/2} (\prod_{k=1}^n \sigma_{n0k}^2)^{1/2}} \exp \left[- \sum_{k=1}^n \frac{(\hat{\tau}_{0k} - t_0)^2}{2\sigma_{n0k}^2} \right] \quad (17)$$

$$\text{将 (16)、(17) 代入 (18) 式:} \quad \left[\frac{\partial \ln P(\hat{\tau} | t_0)}{\partial t_0} + \frac{\partial \ln P(t_0)}{\partial t_0} \right]_{t_0 = \hat{t}_0} = 0 \quad (18)$$

$$\text{得到 } t_0 \text{ 的最大后验概率估计:} \quad \hat{t}_0 = \frac{\sum_{k=1}^n \hat{\tau}_{0k} \sigma_{n0k}^{-2} + \hat{\theta} \sigma_{t_0}^{-2}}{\sum_{k=1}^n \sigma_{n0k}^{-2} + \sigma_{t_0}^{-2}} \quad (19)$$

$$\text{进一步求得:} \quad \sigma_{\hat{t}_0}^2 \triangleq D[\hat{t}_0] = \frac{\sum_{k=1}^n \sigma_{n0k}^{-4} \sigma_{\tau_{0k}}^2 + \sigma_{t_0}^{-4} \sigma_{\hat{\theta}}^2}{\left(\sum_{k=1}^n \sigma_{n0k}^{-2} + \sigma_{t_0}^{-2} \right)^2} \quad (20)$$

态势估计中, 后向推理也很重要。例如: 如果我们获得了某一子任务输出事件发生的时间 $\hat{T}_0 = (\hat{t}_0, \sigma_{\hat{t}_0}^2)$, 但这一时间值含有很大不确定性 ($\sigma_{\hat{t}_0}^2$ 的值很大), 这时需要进一步获取输入数据来减小 $\hat{T}_0$ 的不确

定;为使获取输入数据的代价较小,需要估计传感器最可能得到所需证据的时间。一般情形下,除已知 T_0 外,还已知子任务部分输入事件发生的时间,如 $T_k = (\tau_k, \sigma_k^2) (k = 1, 2, \dots, m; m < n)$ 。后向推理的任务在于估计 $T_j, j = m + 1, \dots, n$ 。因此,首先将 $\hat{T}_0, \hat{T}_k (k = 1, 2, \dots, m)$ 看作 T_0 的 $m + 1$ 个部分预测值,按照第 2 部分的方法对 T_0 进行估计,得到类似 (19)、(20) 的 $\hat{\tau}_0^m, \hat{\sigma}_0^2$ 的表达式,上标号 m ,表示估计值由输入事件中 m 个已知事件发生的时间点估计得到。然后,运用预测关系 $F(t_i, \Delta t_{j0}, \sigma_0^2, t_0)$ 的对称性,对 $j = m + 1, \dots, n$,可得到:

$$\hat{\tau}_j^m = \hat{\tau}_0^m - \Delta t_{j0} \tag{21}$$

$$\sigma_j^2 = \sigma_0^2 + \sigma_{j0}^2 \tag{22}$$

这样,通过基于 MAP 的后向时间推理方法得出对输入事件发生的时间的不确定估计。

2 性能分析和实验结果

这部分以前向推理为例,首先从理论上对基于 MLE 的方法和基于 MAP 的方法的性能进行分析和比较,然后给出实验结果。

2.1 理论分析

从前面的讨论可以知道,采用基于 MLE 的方法对 t_0 进行估计的结果为 $(\hat{\theta}, \hat{\sigma}_0^2)$,如 (13)、(14) 式,采用基于 MAP 的方法对 t_0 进行估计的结果为 $(\hat{\tau}_0, \hat{\sigma}_0^2)$,如 (19)、(20) 式。

(1) 由于 $E[\tau_0] = E[t_0], E[\theta] = t_0$, 所以两种估计都是无偏估计。

(2) 由于 $\forall k \quad M, \sigma_{\tau 0k}^2$ 已知,且 $\sigma_{\tau 0k}^2 = \sigma_{\tau 0}^2 + \sigma_{n0k}^2$,所以 $\sigma_{\tau 0}^2 \rightarrow 0$ 时, $\sigma_{n0k}^2 \rightarrow \sigma_{\tau 0k}^2$,由 (19) 式和 (20) 式,可得 $\lim_{\sigma_{\tau 0}^2 \rightarrow 0} \hat{\sigma}_0^2 = \hat{\sigma}_0^2, \lim_{\sigma_{\tau 0}^2 \rightarrow 0} \hat{\tau}_0 = \hat{\theta}$ 。因此,基于 MLE 的方法是基于 MAP 的方法的一种特例。

(3) 令 $\sigma_{\tau 0l}^2 = \min_k (\sigma_{\tau 0k}^2)$,当 $\sigma_{\tau 0}^2 \rightarrow 0$ 时, $\sigma_{\tau 0}^2 \rightarrow \sigma_{\tau 0l}^2$ 。由 (20) 式可得 $\sigma_{\tau 0}^2 \rightarrow \sigma_{\tau 0l}^2$ 。显然, $\sigma_{\tau 0l}^2 > \frac{1}{\sum_n \sigma_{\tau 0k}^2}$ 。

从 (2), (3) 可以看出,当 $\sigma_{\tau 0}^2 \rightarrow 0$ 和 $\sigma_{\tau 0}^2 \rightarrow \min_k (\sigma_{\tau 0k}^2)$ 时,基于 MAP 的方法不优于基于 MLE 的方法。实验证明,当 $\sigma_{\tau 0}^2 / \min_k (\sigma_{\tau 0k}^2) \leq 0.87$ 且当 $\sigma_{\tau 0}^2 / \max_k (\sigma_{\tau 0k}^2) \geq 0.09$ 时(这一范围我们称之为 $\sigma_{\tau 0}^2$ 的 MAP 适用范围),基于 MAP 的方法优于基于 MLE 的方法。

(4) 当相互独立的部分预测关系的数目 n 增加到 N ,而且所有增加的部分预测值在 $\sigma_{\tau 0}^2$ 的 MPA 适用范围内时,对于基于 MAP 的方法,随着 n 的增大, $\sigma_{\tau 0}^2$ 减小;对于基于 MLE 的方法,由 (14) 式可知,只要 n 增大, $\hat{\sigma}_0^2$ 必然会下降。

2.2 仿真实验结果

实验 1: 令 $n = 12, \sigma_{\tau 01}^2 = 3.2, \forall k \in \{1, 2, \dots, 12\}, \sigma_{\tau 0k}^2 = \sigma_{\tau 01}^2 + 0.2 \times (k - 1)$

当实验次数 i 从 1 到 110 递增时, $\sigma_{\tau 0}^2$ 从 3.0 以步长 0.025 递减。对每个 $\sigma_{\tau 0}^2$ 的取值分别计算 $\hat{\sigma}_0^2$ 和 $\hat{\sigma}_{\tau 0}^2$,得到图 1。图 1 中横轴表示实验次数 i ,纵轴表示估计值的方差,曲线 a 表示 $\hat{\sigma}_0^2$ 随 i 的变化,曲线 b 表示 $\hat{\sigma}_{\tau 0}^2$ 随 i 的变化。从图中可看出在 $i \in (20, 100)$ 时,即 $\sigma_{\tau 0}^2$ 在它的 MAP 适用范围内时, $\hat{\sigma}_{\tau 0}^2 < \hat{\sigma}_0^2$ 。这与理论分析结果 (3) 一致。

实验 2: 当 n 分别为 3, 6, 9, 12 时,对于 n 值, $\sigma_{\tau 0k}^2$ 和 $\sigma_{\tau 0}^2$ 取值如下:

$$\forall k \in \{1, 2, \dots, n\}, \sigma_{\tau 01}^2 = 3.2, \sigma_{\tau 0k}^2 = \sigma_{\tau 01}^2 + 0.2 \times (k - 1)$$

$\sigma_{\tau 0}^2$ 从 3.0 以步长 0.05 递减,这样得到图 2 中的四条曲线和图 3 中的四条曲线。图 2 和图 3 中横轴都表示实验次数 i ,图 2 中的纵轴表示 $\hat{\sigma}_{\tau 0}^2$,图 3 中的纵轴表示 $\hat{\sigma}_0^2$,图 2 与图 3 中的标号“a, b, c, d”依次表示 $n = 3, 6, 9, 12$ 时估计值的方差变化曲线。可以看到随着 n 的增大, $\hat{\sigma}_{\tau 0}^2$ 减小, $\hat{\sigma}_0^2$ 亦减小,这与理论分析结果 (4) 一致。

3 结论

本文首先建立了时间推理方法中已知信息和未知时间变量之间的关系模型,然后将未知时间变量

扩展为一个随机变量，将基于 MAP 的估计方法引入时间推理中，并对 Kirillov 的方法和基于 MAP 的方法的性能进行了分析和比较。理论分析和实验结果都证明：在一个较宽的范围内，基于 MAP 的方法比 Kirillov 的方法性能更好。目前，作者正致力于将本文的方法推广到时间变量为区间的情形，并将统计理论引入 Allen 的理论中，使得 Allen 的理论可以用来处理含有统计不确定性的时间知识。

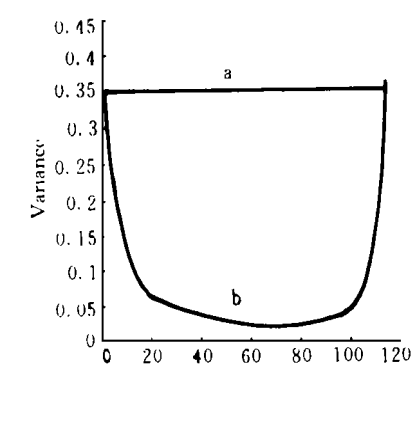


图 1 两种估计方法的估计方差比较

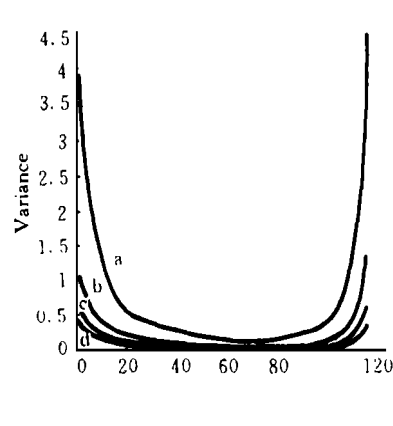


图 2 基于 MAP 的估计值的方差随 n 变化的情况

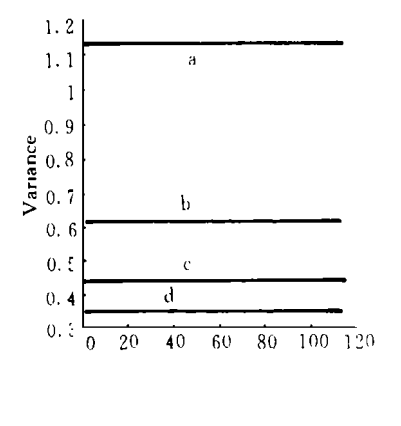


图 3 基于 MLE 的估计值的方差随 n 而变化的情况

参考文献

1 Allen J F. Maintaining Knowledge about Temporal Intervals. Communication of ACM, 1983, 26 (11): 832 ~ 843

2 Dubois D, Prade H. Processing Fuzzy Temporal Knowledge. IEEE Trans. on Systems, Man and Cybernetics, 1989, 19 (4): 729 ~ 744

3 Kirillov V P. Constructive Stochastic Temporal Reasoning in Situation Assessment. IEEE Trans. on Systems, Man and Cybernetics, 1994, 21 (7)