Supplemental materials: Inferring hidden statuses and actions in video by causal reasoning

Amy Fire and Song-Chun Zhu
University of California, Los Angeles
amy.fire@ucla.edu, sczhu@stat.ucla.edu

1 The Viterbi algorithm

In this section, we expand our development of our Viterbi algorithm for the hidden semi-Markov model (HSMM), following the notation and development used in [1]. Let $V_t(pg, \tau)$ be the maximum likelihood that partial state sequence ends at t in state pg of duration τ.

In the paper, we showed the HSMM in Figure 1 below.

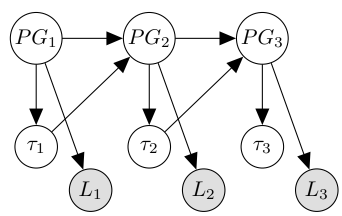

Figure 1: Hidden semi-Markov model

To help develop the Viterbi equations, however, we introduce C_t to indicate that τ_t is complete and, hence, the state is now allowed to change to PG_{t+1} and select a new duration τ_{t+1}.

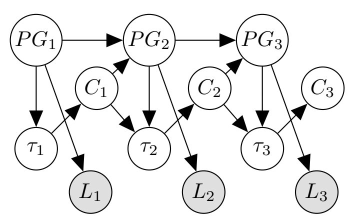

Figure 2: Hidden semi-Markov model with completion nodes

We denote variables as follows:

- PG_t is the random variable for the state space, covering parse graphs from the C-AOG.
- τ_t is the duration that the sequence remains in state PG_t. It is set following $P(\tau|F_t)$ and then deterministically counts down.
- C_t is a binary indicator for whether $PG_t = pg$ is complete and the system is moving to a new pg' and τ'.
- $l_{1:t}$ is the subsequence emitted from 1 to t, and incorporates detections for actions and fluents.
- d is a counter for the current duration countdown

CVPR 2017 Workshop on Vision Meets Cognition

- $1(a, b)$ is the Dirac delta function

Under Figure 2, the hidden semi-Markov model is governed by the following 4 conditional probability distributions:

1. Transition states:
$$P(PG_t = pg|PG_{t-1} = pg', C_{t-1} = c) = \begin{cases} 1(pg, pg'), \text{ if } c = 0 \text{ (remain in same state)} \\ P(pg|pg'), \text{ if } c = 1 \text{ (transition)} \end{cases} \tag{1}$$

2. Reset the duration counter:
$$P(\tau_t = d'|PG_t = pg, C_{t-1} = 1) = P(\tau = d'|F) \tag{2}$$

3. Continue counting down:
$$P(\tau_t = d'|\tau_{t-1} = d, PG_t = pg, C_{t-1} = 0) = \begin{cases} 1(d', d-1), \text{ if } d > 0 \\ \text{undefined, if } d = 0 \end{cases} \tag{3}$$

4. Set to complete when counter is at 0:
$$P(C_t = 1|\tau_t = d) = 1(d, 0) \tag{4}$$

Using C_t, we define:

$$\begin{aligned}
V_t(pg, \tau) &\triangleq \max_{pg', \tau'} P(PG_t = pg, C_t = 1, \tau_t = \tau, \\
&\qquad PG_{t-1} = pg', \tau_{t-1} = \tau', C_{t-1} = 1, l_{1:t}) && (5) \\
&= \max_{pg', \tau'} [P(l_{t-\tau+1:t}|PG_t = pg) && (6) \\
&\qquad P(PG_t = pg, \tau_t = \tau, PG_{t-1} = pg', \tau_{t-1} = \tau, C_{t-1} = 1, l_{1:t-\tau})] \\
&= \max_{pg', \tau'} [P(l_{t-\tau+1:t}|PG_t = pg) && (7) \\
&\qquad P(PG_t = pg, \tau_t = \tau|PG_{t-1} = pg', \tau_{t-1} = \tau) \\
&\qquad P(PG_{t-1} = pg', \tau_{t-1} = \tau', C_{t-1} = 1, l_{1:t-\tau})] \\
&= P(l_{t-\tau+1:t}|PG_t = pg) && (8) \\
&\qquad \max_{pg', \tau'} [P(PG_t = pg, \tau_t = \tau|PG_{t-1} = pg') \\
&\qquad P(PG_{t-1} = pg', \tau_{t-1} = \tau', C_{t-1} = 1, l_{1:t-\tau})] \\
&= P(l_{t-\tau+1:t}|PG_t = pg) \\
&\qquad \max_{pg', \tau'} P(PG_t = pg, \tau_t = \tau|PG_{t-1} = pg')V_{t-\tau}(pg', \tau') && (9) \\
&= P(l_{t-\tau+1:t}|pg) \max_{pg', \tau'} P(pg, \tau|pg')V_{t-\tau}(pg', \tau') && (10)
\end{aligned}$$

Since we assume the conditional independence

$$P(pg, \tau|pg') = P(pg|pg')P(\tau|pg, pg') = P(pg|pg')P(\tau|pg) = P(pg|pg')P(\tau|F), \tag{11}$$

$V_t(pg, \tau)$ becomes

$$\begin{aligned}
V_t(pg, \tau) &= P(l_{t-\tau+1:t}|pg) \max_{pg', \tau'} P(pg|pg')P(\tau|F)V_{t-\tau}(pg', \tau') && (12) \\
&= P(l_{t-\tau+1:t}|pg)P(\tau|F) \max_{pg', \tau'} P(pg|pg')V_{t-\tau}(pg', \tau') && (13) \\
&= P(l_{t-\tau+1:t}|pg)P(\tau|F) \max_{pg'} \left[P(pg|pg') \max_{\tau'} V_{t-\tau}(pg', \tau')\right]. && (14) \\
& && (15)
\end{aligned}$$

To separate the duration from the state space, define:

$$V_t(pg) \triangleq \max_{\tau} V_t(pg, \tau). \tag{16}$$

Therefore,

$$V_t(pg) = \max_{\tau} \left[P(l_{t-\tau+1:t}|pg)P(\tau|F) \max_{pg'} [P(pg|pg')V_{t-\tau}(pg')]\right]. \tag{17}$$

References

[1] K. Murphy. Hidden semi-markov models (hsmms). Unpublished notes, 2002.