Chapter 6

Behavior is Reward-oriented

...

6.1 Reinforcement learning

...

6.1.1 Speeding up temporal difference learning

...

Eligibility traces

...

However, when focusing on temporal difference learning, the agent does not have a model about its environment and thus cannot look forward in time. However, the same principle also works backwards in time. In this case we need to maintain a memory of previously encountered states and of the executed actions in these previous states. During each update then, not only the current state-value or state-action value is updated, but also all remembered previous ones. Typically though, not all previous ones should have the same update strength, but more recent states should undergo stronger updates. This is accomplished by determining an *eligibility* of each previous state. The eligibility is easy to determine when defining it using the most recent point in time a particular state had been visited:

$$e_t(s) \quad = \quad \begin{cases} (1-\lambda)(\lambda\gamma)^{t-k} & \texttt{if } k \geq 0 \\ 0 & \texttt{otherwise} \end{cases}, \tag{6.1}$$

where $k = -1$ if state s has not been visited at all so far, and $k = \max\{k|s_k = s\}$, otherwise. The factor λ determines the spread of the eligibility, where $\lambda = 0$ corresponds to a normal TD update, while $\lambda \to 1$ spreads the eligibility uniformly into the past. It is guaranteed that not too much reward is distributed into the past, as $\sum_{t'=0}^{t} e_t(s_t) \leq 1$, because $e_t(s)$ is bound by the geometric series: $\sum_{x=0}^{\infty}(\gamma\lambda)^x = \frac{1}{1-\gamma\lambda} \leq \frac{1}{1-\lambda}$ (given $0 \leq \lambda < 1$ and $0 < \gamma \leq 1$) .

With the concept of eligibility, the temporal difference update is applied to all states that have been encountered so far until time t, yielding the enhanced temporal difference update equation:

$$V^\pi(s) \quad \leftarrow V^\pi(s) + \alpha \cdot e_t(s) \cdot [R_{t+1} + \gamma \cdot V^\pi(s_{t+1}) - V^\pi(s_t)]\,, \tag{6.2}$$

SARSA learning can be updated accordingly. Q-learning, on the other hand, is not directly applicable because the trace depends on the policy, violating the off-policy principle of Q-learning updates. Nonetheless, initial faster learning can also be achieved in this case, while full convergence to $Q^\star$ relies on proper, off-policy Q-value estimation updates.

...

Chapter 11

Multisensory Interactions

...

11.1 Body-relative spaces

...

11.1.1 Optimal information fusion

...

Equipped with these measurements it can be shown that the optimal sensory integration is accomplished by the following equation, when assuming that the information sources, which are integrated in a particular location space L, are independent of each other:

$$\hat{L}(t) = \frac{\sum_{i \in \mathcal{I}} f_i(s_i(t)) \frac{1}{f_i(\sigma_i^2(t))}}{\sum_{j \in \mathcal{I}} \frac{1}{f_j(\sigma_j^2(t))}}, \tag{11.1}$$

where the denominator is a constant that essentially normalizes the estimate, yielding a proper relative precision-weighted integration of location estimates, and $\mathcal{I}$ denotes the set of sensory information sources that contribute to the location estimate. The equation essentially multiplies $|\mathcal{I}|$ independent information source estimates.

The resulting variance estimates can be calculated by:

$$\hat{\sigma}(t) = (\sum_{j \in \mathcal{I}} \frac{1}{f_j(\sigma_j^2(t))})^{-1}, \tag{11.2}$$

again assuming information source independence.

The resulting estimates correspond to the *maximum likelihood estimate* of information theory. That is, $\hat{L}(t)$ is the maximum likely location when assuming that all information sources about the location are independent and the respective uncertainties σ_i^2 can be projected without biases into the location space. While these assumptions are not totally valid in most cases, the estimate typically serves as a good approximation. Note how this estimate is strongly related to Gaussian distributions: Eq.(11.1) and Eq.(11.2) can indeed be calculated in closed-form when all individual distributions and their respective projections into the location space are Gaussian (Rasmussen & Williams, 2006), with means and variances in location space specified by $f_i(s_i(t))$ and $f_i(\sigma_i^2(t))$, respectively.

...

... Given that a motor command was executed, we then encounter an actual spatial transition, which yields an estimate about the resulting location:

$$\hat{L}'(t+1) = \hat{L}(t) + g(m(t)). \tag{11.3}$$

Note how this estimate is related to two concepts, which were introduced in previous chapters: first, we have formalized the *reafference principle* (cf. Section 6.4.1), which anticipates the sensory consequences – in this case actually the location consequences – given a motor command; second, we have generated an *a priori estimate* of a location, given information from the past, according to Bayesian information processing principles (cf. Section 9.3).

To expand the location anticipation to a full probability density estimate, we require some sort of uncertainty estimate in order to yield a location-distribution, rather than one location estimate. Again, let us keep things simple and assume that a variance estimate $\sigma_L^2(t)$ is carried along. How should this estimate change over time? Assuming that we have an a priori location estimate in the form $[\hat{L}'(t), \sigma_L'^2(t)]$ available, we may then consider the incoming sensory information. Assuming further that the location estimate itself is independent of all sensory information (which is typically not the case but is assumed here to keep things still reasonably simple; cf. Kneissler, Drugowitsch, Friston, & Butz, 2015 for a more exact derivation, which avoids falling into self-fulfilling delusions when the internal estimates are overly trusted) , the location estimate can be considered as a fully independent estimate, similar to the sensor-based location estimates. Thus, information fusion simply extends to:

$$\hat{L}(t) = \frac{\left(\sum_{i\in\mathcal{I}} f_i(s_i(t))\frac{1}{f_i(\sigma_i^2(t))}\right) + \hat{L}'(t)\frac{1}{\sigma_L'^2(t)}}{\sum_{j\in\mathcal{I}}\frac{1}{f_j(\sigma_j^2(t))} + \frac{1}{\sigma_L'^2(t)}}, \tag{11.4}$$

and, for the variance estimate, to:

$$\hat{\sigma}_L^2(t) = \left(\sum_{j\in\mathcal{I}}\frac{1}{f_j(\sigma_j^2(t))} + \frac{1}{\sigma_L'^2(t)}\right)^{-1}, \tag{11.5}$$

yielding the a posteriori location estimate, where the a posteriori uncertainty mixes the a priori uncertainty with other independent information sources, yielding information gain, that is, a decrease in uncertainty.

When then projecting the location estimate into the future by means of the motor-dependent projection function $g(m(t))$, the location will be shifted and the uncertainty should again increase to a certain extent. This extent may depend on the motor function, but it may also add by default some uncertainty, such that, for example:

$$\hat{L}'(t+1) = \hat{L}(t) + g(m(t)), \tag{11.6}$$

$$\sigma_L'^2(t+1) = \hat{\sigma^2}_L(t) + g(\sigma_m^2(t)) + \sigma_c^2, \tag{11.7}$$

where σ_c^2 adds uncertainty, which may account, for example, for neural processing noise. As a result, the processing loop is closed and the system can continuously maintain an internal estimate $[\hat{L}'(t), \sigma_L'^2(t)]$.

The formalized loop certainly simplifies the actual neurocognitive processing that is going on in several respects. Moreover, it is not known to what extent and exactly how the put-forward formalization is implemented by the brain. However, from a computational perspective, some sort of processing, which mimics this optimal information processing sketch, needs to be realized in order to be able to maintain internal spatial estimates about locations in the environment, as well as about one's own body posture. Various behavioral studies have confirmed that a process similar to this one is at work (Butz, Kutter, & Lorenz, 2014; Ehrenfeld, Herbort, & Butz, 2013; Ernst & Banks, 2002).

Advanced formalizations of these equations can be derived from free-energy-based minimization principles, thus providing an even more general formalization (Friston, 2009; Kneissler et al., 2015). Additional information processing steps appear to be at work. In particular, it appears that different sensory information sources are compared with each other, fusing only those sensory information sources with the a priori location estimates that provide *plausible* information. Moreover, the resulting a posteriori spatial estimates

may be further compared with other information sources – such as estimates about other objects – further fostering the consistency between these estimates given knowledge about the body and the environment. For example, an object may not be exactly located where another object is already located. Similarly, the limbs of the body can only be arranged in certain ways, given limb lengths and joint angle flexibilities. In fact, experimental and modeling results suggest that our brain attempts to maintain a consistent postural body schema estimate over time (Butz et al., 2014; Butz, 2016; Ehrenfeld et al., 2013).

...

References

Butz, M. V. (2016). Towards a unified sub-symbolic computational theory of cognition. *Frontiers in Psychology*, *7*(925). doi: 10.3389/fpsyg.2016.00925

Butz, M. V., Kutter, E. F., & Lorenz, C. (2014). Rubber hand illusion affects joint angle perception. *PLoS ONE*, *9*(3), e92854. doi: 10.1371/journal.pone.0092854

Ehrenfeld, S., Herbort, O., & Butz, M. V. (2013). Modular neuron-based body estimation: Maintaining consistency over different limbs, modalities, and frames of reference. *Frontiers in Computational Neuroscience*, *7*(148). doi: 10.3389/fncom.2013.00148

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293 - 301. doi: 10.1016/j.tics.2009.04.005

Kneissler, J., Drugowitsch, J., Friston, K., & Butz, M. V. (2015). Simultaneous learning and filtering without delusions: a bayes-optimal combination of predictive inference and adaptive filtering. *Frontiers in Computational Neuroscience*, *9*(47). doi: 10.3389/fncom.2015.00047

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.