Anticipation in Cognitive Radar Using Stochastic Control

Alexander Charlish
Fraunhofer FKIE
Fraunhoferstr. 20
53343, Wachtberg
GERMANY

alexander.charlish@fkie.fraunhofer.de

ABSTRACT

This paper introduces a cognitive radar system concept, where multiple cognitive processes are manifested by a system comprised of potentially non-cognitive components. Then, examples of techniques developed over the last decades that possess cognitive processes are given. This is followed by a description of currently underdeveloped cognitive processes from the perspective of cognitive psychology. The paper then focuses on the cognitive process of anticipation, which can be generated using partially observable Markov decision processes (POMDPs). An introduction to POMDPs is given followed by a description of algorithms that can give approximated solutions. Finally, an anticipative target tracking example is given, where the benefit of anticipation is demonstrated in comparison to standard adaptive tracking techniques.

1.0 COGNITIVE RADAR CONCEPT

Cognitive radar can be defined as a radar system that possesses the cognitive processes that are required for cognition. The definition of what cognitive processes are can vary, however, standard texts on cognitive psychology [1], [2] all list very similar cognitive processes, These are perceptual processes (perception generation, attention, recognition), memory processes (long term memory, working memory, learning), language processes (concepts and categorisation, language processing, language comprehension and language production) as well as thinking processes (problem solving, reasoning, judgement, decision making and anticipation). It is evident from the varied cognitive development of animal species that cognitive processes can be very basic or highly sophisticated, resulting in a long ladder with many levels of cognitive development. A key question for the topic of cognitive radar is to determine where on this ladder the border between the radar system and the operator lies. The topic of cognitive radar ultimately aims to shift the responsibility of the radar system in performing cognitive processes further up this ladder, alleviating the operator's cognitive responsibilities.

1.1 Existing Techniques

The concept of cognitive radar was born out of a long string of technological developments. In fact, it is difficult to think of any radar technological development that does not contribute to one of the cognitive process listed previously. The partial subset of perception, action and memory has been highly prevalent in radar resource management research over the last decades. This section gives an overview of some of these techniques before describing the cognitive processes that are currently underdeveloped in the context of cognitive radar.

1.1.1 Radar Search Management

For radar search the cognitive process of attention is important, so that the search is focused on important regions of the current situation. The radar search function traditionally executes a fixed search pattern such as a raster that can be defined by the waveform and the time taken to complete the search, which is particularly non-cognitive. As a first step in achieving attention, a search volume can be divided into a number of sectors with varied frame times, waveforms and hence resource allocations, to reflect different sensing requirements dictated by the situation [3]. To further focus the attention of the search to the current situation, an undetected target density [4], [5] can be utilised, which is a perception of the current state of the search problem. Search dwells can then be planned to maximise the probability of detecting previously undetected targets in the current situation. Executing search measurements based on an undetected target density constitutes a perception-action cycle with memory and decision-making.

In addition to a perception of undetected targets, search radars are also controlled based on a perception of the environment in which they operate. Clutter maps describing the perception of the spatial distribution of clutter can be generated and exploited for adaptive signal processing. Also, jammer assessment can be performed enabling a perception of the intentional interference to be generated. This perception can be used for controlling the search, for example by switching to an interference free frequency channel. Finally, search radar perceives its own health and performance, for example through calibration or detecting antenna element failures. This self-performance perception is the basis for self-healing or for planning repairs.

1.1.2 Track Management

In comparison to search management, radar track management has received significantly more attention. Adaptive tracking techniques [6] enable the revisit interval time between measurements to be varied depending on target manoeuvres, such that the target track is maintained with the minimum radar resource loading. The key to the approach is to schedule measurements when the estimation error standard deviation of the target's angular position reaches a fraction of the 3dB radar beamwidth. Additionally, the target radar cross section can be estimated, so that a waveform with the shortest dwell length that achieves a specified signal-to-noise ratio can be used. Through the benchmark problems [7]–[9], the importance of interacting multiple model filtering was demonstrated, as it ensures that the target dynamic model in the tracking filter is matched to the current target manoeuvre, hence enabling adaptive tracking. The benchmark problems also demonstrated that the performance of the perception process (tracker) strongly influences the resulting action (resource management), which highlighted the complementarity between perception (assessment) and action (management) processes. Despite possessing a perception–action cycle with memory, these techniques have only been described as 'adaptive' over the last two decades.

In addition to controlling the revisit interval time and the desired signal-to-noise ratio, the intra-pulse modulation can also be selected based on the current target track (perception). Consequently, waveforms with shorter time-bandwidth products or traditionally `bad' ambiguity properties may be preferred, when they fulfil the specific information need of the tracker. Kershaw and Evans [10], gave closed form solutions for selecting linear FM chirp waveforms that minimise the tracking mean squared error, and later extended the approach to include clutter [11]. Although arbitrary waveform generation is now possible, the online design of waveforms can be real-time infeasible, therefore the use of small but well-designed waveform libraries has been proposed [12]. Considering waveform selection from the track performance perspective also motivates the use of non-traditional waveforms, such as non-linear FM waveforms [13] or fractional Fourier transformed waveforms [12]. Haykin [14], [15] took a similar approach, unifying waveform selection with a concept of cognitive radar. As with search management, track management can also be based on a perception of the current clutter and jamming environment.

1.1.3 Matched Illumination

At the signal level, the transmit waveform and receiver filter can be managed based on knowledge of the radar channel [16]. Through the KASSPER program [17] it was shown that knowledge of the multi-dimensional radar channel enables the transmit waveform to be matched to the estimated channel model. At the receiver side, space, time, or space and time adaptive processing can be applied based on knowledge of the interference. As the radar channel and interference can be dynamic and uncertain, knowledge must be learnt online based on observed radar data. The efficacy of this learning process depends greatly on the availability of training data. This waveform adaptation approach differs to the waveform selection methods in the track management section due to the abstraction level at which it is applied. The waveform selection methods based on track information aim to maximise object level performance criteria, such as the track estimation error. In contrast, these waveform adaptation methods aim to maximise a signal level criterion, which is the signal-to-interference ratio. However, both approaches control the waveform based on a perception–action cycle.

1.1.4 Quality-of-Service Optimisation

Quality-of-Service (QoS) management methods [18] enable task control parameters to be selected for multiple tasks, by considering the trade-off between each task's utility contribution and resource usage. The task utility contribution is determined by a mapping from a task relevant quality measure into utility, and represents the mission relevant satisfaction associated with the task quality level. QoS management methods incorporate perception, action and memory. As QoS management methods optimise the balance of the resource allocation between the competing tasks, they also enable the process of attention. Effective resource management using QoS optimisation methods is the topic of a following paper in this lecture series.

1.1.5 Stochastic Control

Stochastic control is method that encompasses perception, memory, action, decision making and potentially also learning and has been applied to radar problems. Stochastic control applied to radar management problems is the topic of this paper.

1.2 Cognitive Processes

From the examples listed previously, it can be seen that perception action cycles with memory have been present in radar research over a number of decades, mostly described as adaptive approaches. In fact, radars possessing limited cognitive processes are already operational [3], [19], [20]. However, many cognitive processes are underdeveloped in the context of cognitive radar. This section describes these underdeveloped cognitive processes.

Learning

Learning is the process of acquiring new knowledge on the environment, which is used to enhance perception generation, as well as to take well-informed decisions and execute well-informed actions. The relevance of learning for cognitive radar can be seen, for example, in matched illumination where it is necessary to learn a model of the radar channel in order to control the degrees of freedom for the transmit waveform and the receiver filter. The process of learning must be a defining feature for cognitive radar.

Problem Solving

The process of problem solving is critical to cognition. For cognitive radar that should take on the cognitive processes of a human operator, the goal is to satisfy the mission requirements and objectives of the operator. The radar management methods listed in Sec. 1.1 formulate objective functions based on lower level performance criteria, such as the signal-to-interference ratio or track estimation error. It is then implicitly assumed that optimising these lower level criteria aggregate to a successful mission. However, a successful mission may not be achieved as the true goal is not explicitly considered. Therefore, cognitive radar should exhibit goal-orientated behaviour that is focused on mission objectives.

Concepts and Categories

A concept is an internal idea that applies to a category of things, enabling sets of objects to be sorted into categories. Categorising according to a concept is a key human cognitive process that enables humans to respond to objects depending on category instead of the unique object itself. The cognitive process of categorisation based on concepts may sound abstract, however, it bears a striking resemblance to higher-level information fusions systems and situation assessment methods [21]. A cognitive system should comprehend the situation it is in and act accordingly, hence cognitive radar should exploit situation assessment and management methods.

Language

For a cognitive radar system it is necessary for the system to communicate effectively with the operator through the human machine interface (HMI). Not only must the operator be able to effectively communicate objectives and requirements, the radar system must provide the necessary information to justify the decisions that the radar system takes, otherwise the operator will not trust the radar. In addition to the operator, it is desirable for the radar system to communicate with other sensor systems or platforms. Developing networked sensor systems is currently a large and evolving area of research.

Judgement, Decision Making and Reasoning

Radar management is relatively underdeveloped in comparison to adaptive processing at the receiver. Therefore, cognitive radar must advance the decision-making processes applied in radar management. Reasoning is the process of inferring a conclusion based on premises, by following logical laws. Categorisation plays a crucial role in reasoning, for example, when someone drives a previously unfamiliar car, although nothing is known about the car it is possible to reason that the car has brakes due to the person's concept of a car.

1.3 System Architecture

In order to manifest many cognitive processes, a system view can be taken where certain cognitive processes are spread over components in the system architecture. Each system component may not be described as cognitive in isolation but can contribute to the cognitive behaviour of the complete system.

A cognitive radar system architecture can be constructed based on a hierarchy of information abstraction levels, where perception, action and memory are located at each level. Abstraction levels for sensor data and information processing have been widely studied, most notably by the JDL model [22] and its revised versions [23], [24]. Based on these studies, information abstraction levels relevant for a cognitive radar system can be identified as the signal, burst, measurement, object, situation and mission levels. This information abstraction hierarchy acts as a bridge between radar signals and the operator's mission objectives or requirements. Fig. 1 illustrates an adapted version of the cognitive radar architecture by Kester [25] and Smits et al. [26]. Although the concept of a fully cognitive radar system is in its infancy, it can be seen that many of the components that can contribute to system cognition not only pre-date the concept of cognitive radar but are also technologically mature. In fact, it is difficult to identify any radar technological development that could not be argued as contributing in some way to one of the cognitive processes listed previously.

The cognitive processes discussed here were heavily based on human cognition, however, a radar system will never truly act like a human. Therefore it is necessary to question what level of cognitive development is actually required and where on the ladder of cognitive development the border between the radar system and the operator lies. Also, cognitive radar would not come without any drawbacks, such as unpredictability and potential exposure to new forms of electronic attack.

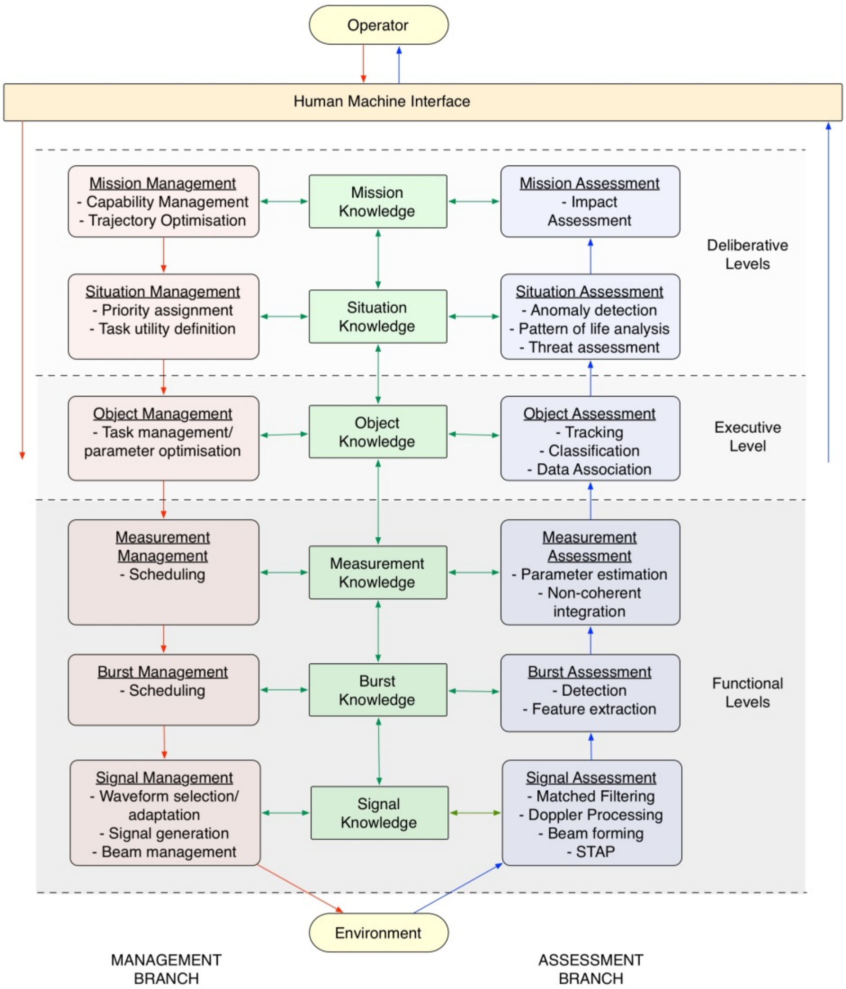

Figure 1-1: Cognitive Radar System Architecture.
(Cognitive radar system architecture comprised of a number of non-cognitive components. The architecture is structured in information abstraction levels with perception, action and learning performed at each level. The operator interacts with the radar through the HMI, which accesses assessment data and control parameters.)

2.0 STOCHASTIC CONTROL

The remainder of this talk focuses on the process of anticipation using stochastic control. This section presents a brief overview of a POMDP problem formulation. In depth handling of POMDPs can be found in many places in the literature, and this overview follows the formulation by Chong et al. [27].

2.1 Partially Observable Markov Decision Process

A POMDP is a sequential decision making process where actions are sought at each decision instance, which maximise rewards that are accumulated over a time horizon in the future. A POMDP consists of the following components:

State Space - The state space X describes the range of possible states of the system, where a state at time step k is denoted x_k. For radar tracking the state can be the true positions of the target and the radar platform. For search the state can be the location of undetected targets.

Action Space - The action space A describes the range of possible actions that can be taken, where an action at time step k is denoted a_k. The action can be the scheduling of a measurement at a certain time with a corresponding waveform.

State Transition Probability - The state transition probability function $p(x_{k+1}|x_k, a_k)$ gives the probability of transitioning to state x_{k+1} from state x_k when taking action a_k.

Observation Space - The observation space Z describes the range of possible measurements that can be observed, where a measurement at time k is denoted z_k.

Observation Likelihood Function - The observation or measurement likelihood function describes the probability $p(z_k|x_k)$ of observing measurement z_k given that the system is in state x_k.

Reward Function - The reward function $r(x_k, a_k)$ gives the reward received when action a_k is taken when the system is in state x_k. This reward must reflect the radar's sensing objective.

Given a reward is received for the pairing of the true system state and an action, the objective of the POMDP is to maximise the cumulated reward V_H starting from the time step k_0 up to the end of the time horizon:

$$V_H = E\left[\sum_{k=k_0}^{k_0+H} r(x_k, a_k)\right] \tag{1}$$

At each decision stage the controller is required to select an action a_k, as the first step in the action trajectory that maximises Eq. (1). The selection of an action at decision step k is based on the set of actions that have been performed and the measurements that have been observed prior to time step k, which is denoted as the data set $d_k = \{z_0, a_0, \dots, z_{k-1}, a_{k-1}, z_k\}$.

As the true state of the system is unobservable, the controller forms a belief b of the unobservable system state, which is represented by a probability distribution on the state space X. The belief state b_k is conditioned on the prior data:

$$b_k = p(x_k|d_k) \tag{2}$$

The optimal policy function is then sought, which maps the belief on the system state that the controller currents holds, into the best action to take:

$$\pi^*: b_k \mapsto A \tag{3}$$

A POMDP is illustrated in Fig. 2-1. As shown in the figure, the true state of the system is only partially observable through noisy observations/measurement. Consequently, the controller maintains a belief state b which is used to select the action a which maximises the expected reward over multiple future decision stages up to the specified time horizon H.

In order to make decisions about the best action to perform, it is necessary to evaluate the expected value of a belief state, which is based on the expected reward of possible future system states. As the controller has an uncertain belief on the true system state, let $R(b_k, a_k)$ be the expected reward with respect to belief state b_k. Then, the expected value of belief b_t for a POMDP starting at k_0 is:

$$V_H(b_t) = E\left[\sum_{k=k_0+t}^{k_0+t+H} R(b_k, a_k)|b_t\right] \quad (4)$$

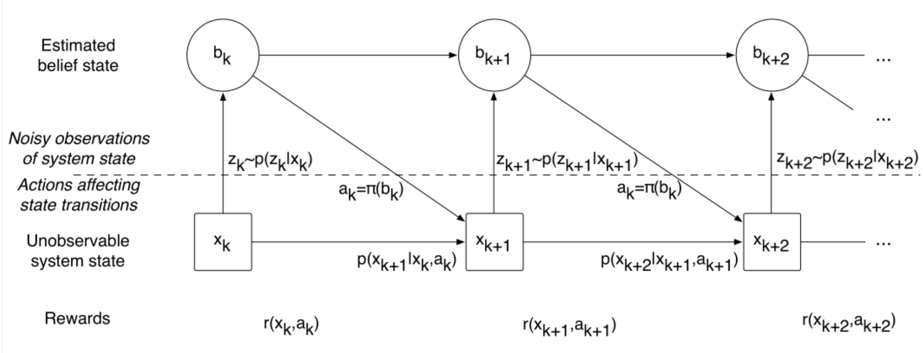

Figure 2-1: Partially observable Markov decision process.
(Illustration of a partially observable Markov decision process. A system state x, which is partially observed through noisy measurements z, is controlled by actions a that trigger Markovian transitions. Rewards are accumulated over the decision stages depending on the system state and the action taken.)

Using Bellman's equation, the optimal value $V_H^*(b_0)$ of the initial belief state b_0 can be decomposed into the expected reward $R(b_0, a)$ of taking action a from belief state b_0 and the optimal value of the subsequent belief state b_1 that would be reached after the action:

$$V_H^*(b_0) = \max_a (R(b_0, a) + E[V_{H-1}^*(b_1)|b_0, a]) \quad (5)$$

based on this, the optimal policy can be defined as the selection of the action that maximises the value of being in the initial belief state:

$$\pi_0^*(b_0) = \arg\max_a (R(b_0, a) + E[V_{H-1}^*(b_1)|b_0, a]) \quad (6)$$

Within this equation, the commonly termed Q-value can be defined as:

$$Q_{H-t}(b_t, a) = R(b_t, a) + E[V_{H-t-1}^*(b_{t+1})|b_t, a] \quad (7)$$

Using the definition of the Q-value, the optimal policy from Eq. (6) can then be rewritten as finding the action that maximises the Q-value:

$$\pi_t^*(b_t) = \arg\max_a Q_{H-t}(b_t, a) \quad (8)$$

To find the optimal action to take at time step t, it is necessary to evaluate the Q-value for all possible candidate actions. The Q-value is comprised of two terms, the instantaneous reward and the possible future reward.

2.2 POMDP Cognitive Processes

The POMDP formulation incorporates the following cognitive processes:

Memory and Perception - The concept of memory and perception is central to the POMDP, as the belief state b_k represents the interpretation of the partially observable system state. This perception is clearly based on memory, as b_k is conditioned on the entire action-measurement history $d_k = \{z_0, a_0, \ldots, z_{k-1}, a_{k-1}\}$.

Actions - Action selection is the core task of a POMDP. The best action is sought based on the memory of previous actions and measurements, and the perception of the partially observable system state.

Anticipation - By evaluating the expect rewards over a future time horizon, a POMDP selects actions based on how the system state is anticipated to evolve in the future. The following two cases demonstrate the differentiation between adaptation and anticipation:

- **Case 1 - Time horizon** $H = 1$: Eq. (1) simplifies to the reward $r(x_k, a_k)$ and therefore the optimal action is based only on the belief b_k of the system state at the current time k. Action selection based only on the current belief state can be thought of as *adaptive*.
- **Case 2 - Time horizon** $H >> 1$**:** Eq. (1) comprises of a trajectory of future actions and states, therefore the POMDP reasons about the rewards it anticipates to receive in the future. This anticipation of future rewards can be considered a *cognitive* process.

The effect of the time horizon in a POMDP is widely discussed in the sensor management literature as myopic (considering only the present) or non-myopic (considering also the future) management.

2.3 Approximate Solutions

Unfortunately, an optimal solution to a POMDP is intractable for all problems except those involving a small number of finite system states [28]. Therefore a lot of research has been dedicated to generating approximate solutions to POMDPs.

2.3.1 Algorithm Types

Solution methods for POMDPs can be separated into offline and online algorithms. Offline algorithms precompute policies for possible belief states before deployment, whereas online algorithms compute policies online based on the current belief of the system state.

Offline Algorithms

For offline algorithms, an action is specified for each belief state that could be encountered. These algorithms rely on the fact that the optimal value function over the belief state is piecewise linear convex [29], and therefore representable with a finite set of vectors, so called α-vectors. A key approach is to use point POMDP solvers [30], [31]. Most of the existing research in offline algorithms is based on discrete systems. However, as radar systems observe a continuous state space from continuous measurements, these methods require an additional discretisation step. Existing algorithms for continuous states and measurements [32], [33], [34], [35] are quite computationally intensive and currently do not scale well for sensor management problems.

Online Algorithms

In comparison to offline algorithms, online algorithms compute policies during deployment. Consequently, it is only necessary to explore belief states that are reachable from the current system belief state. The belief states which follow the current belief state build a tree, where the nodes of the tree are the possible future beliefs, connected by the possible observations and actions. Online algorithms search this tree to effectively approximate the Q-value in Eq. (7).

If the measurements were discrete and finite, the tree could theoretically be exhaustively searched. However for radar, like in most sensor management applications, measurements are considered as continuous and therefore an exhaustive search is impossible. Instead, measurements can be stochastically sampled, deterministically sampled or only the most likely measurement can be considered. Regardless, the tree is typically too big to allow a complete search and therefore several approximation methods have been developed:

Pruning - If it is possible to compute upper and lower bounds for the future reward of a belief state, several branches of the belief tree can be completely ignored, if they cannot contain the optimal future decisions.

Rollout - The rollout method [36] assumes that the controller behaves in the future according to a so-called base policy. Therefore not every future action has to be evaluated, but instead only those actions that are generated from the base policy. The rollout method is described in detail in Sec. 2.3.2.

Approximation of the value of a belief state - In some problems it is possible to approximate the value of a belief state, without the need to further explore the belief tree. In this case the tree only needs to be computed for the first level, and the approximation can be used for the value of the belief.

Reward substitution - Sometimes the reward is hard to compute for the online algorithm, and can instead be replaced by another reward, which captures the same behaviour. For example, often in sensor management the goal is to reduce the RMSE of a target estimate, instead the achieved Fisher Information can be used as a reward.

A detailed overview about online POMDP solutions can be found in [37]. Approximate POMDP solutions in sensor management are covered in [38].

2.3.2 Policy Rollout

According to equation (7) the Q-value consists of two values: The immediate reward and the expected reward in the future. Rollout replaces the expected future value that would be achieved if the optimal policy were followed, with the expected value when a base policy is followed. The base policy is a hand crafted policy which describes a sensible heuristic to generate actions based on an encountered belief state. The rollout procedure traverses the belief tree, while selecting the actions according to the base policy. This process of rollout is motivated by the fact that it is not necessary to calculate the values of each candidate action exactly, it is just sufficient to know the relative rankings of the candidate actions, for the best action to be taken.

Given a base policy:

$$\pi_B : b \rightarrow a \tag{9}$$

the Q-value is replaced by:

$$Q_{H-t}^{\pi_B}(b_t, a) = R(b_t, a) + E[V_{H-t-1}^{\pi_B}(b_{t+1})|b_t, a] \tag{10}$$

where $V_{H-t-1}^{\pi_B}(b_{t+1})$ is the value of belief b_{t+1} if the system follows the base policy π_B. As the optimal value $V_{H-t-1}^{*}(b_{t+1})$ is defined as the value achieved if the controller follows an optimal policy, $V_{H-t-1}^{\pi_B}$ is a lower bound to the optimal value. This value can be computed for example via Monte Carlo simulation of future belief states. Computation can be reduced by simplifying the rollout step, for example, by using the expected measurement instead of multiple Monte Carlo runs with sampled measurements. In this case, care must be taken that the simplified rollout still accurately reflects the trade-offs between the different actions. Policy rollout can be extended to parallel rollout, which uses multiple base policies [39].

3.0 RADAR ANTICIPATION WITH POMDPS

3.1 POMDPS in Radar Applications

POMDPs have been applied in sensor management for `active sensing' [27], [38], [40], [41]. These techniques are applicable for active sensing in general and not just for radar applications, however, a radar model is frequently adopted. A Markov decision process (MDP) approach has been applied to radar problems by Wintenby [42], [43], whereby the partial observability is modelled by a number of discrete states in an MDP. POMDP approaches have also been applied for alternative sensor management problems, such as path planning for a UAV with a radar [44]–[46], and waveform scheduling [47]–[49].

3.2 Anticipative Target Tracking Example

In this section, an example of stochastic control applied to a target tracking problem is described [50].The objective is for the controller to select the time interval between radar measurements for a target track, such that a desired estimation error is achieved with the minimum resource usage. An electronically steered array antenna is assumed, such that measurements are made by steering the beam to the estimated target position. As a scenario may dictate that measurements provide different amounts of information, the anticipated future development of the situation must be taken into account. This is done with a rollout based approach.

3.2.1 Scenario Description

The scenario consists of an airborne radar platform and a Swerling 1 target with nearly constant velocity motion at 200m/s, as illustrated in Fig. 3-1. In the scenario, the target is unobservable during a certain period of time. This non-observability could be due to a number of reasons, such as a blockage to the line of sight, a jammer, or the unavailability of multifunction radar when a different non-interruptible function is executed. It is assumed that the borders of the unobservable region are known.

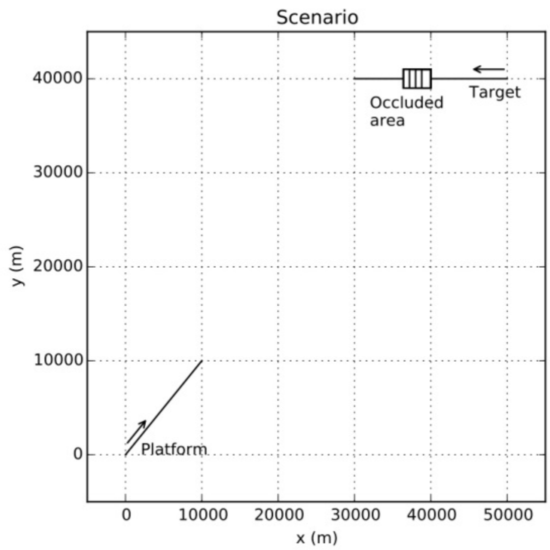

Figure 3-1: Simulated Scenario.
(The target is non-observable for a period of time.)

3.2.2 POMDP Formulation

The problem can be formalized as a POMDP, based on the definition given in Sec. 2.1.

System State

The system state is a stacked vector comprising of the target and platform kinematics $x = (x^p, x^t) \in X \subset \mathbb{R}^{12}$ where $x^p = (x^p, \dot{x}^p, y^p, \dot{y}^p, z^p, \dot{z}^p)^T \in \mathbb{R}^6$ is the position and velocity of the platform, and $x^t = (x^t, \dot{x}^t, y^t, \dot{y}^t, z^t, \dot{z}^t)^T \in \mathbb{R}^6$ is the position and velocity of the target and T is the transpose operator.

A belief on the system state is estimated by the controller using a Kalman filter. Consequently, the belief state b_k at time k is represented by a Gaussian $b_k = N(x_k^t; x_{k|k}^t, P_{k|k})$. The posterior filter state estimate is denoted $x_{k|k}^t$ and the covariance matrix $P_{k|k}$ is the filter calculated MSE in the estimate:

$$P_{k|k} = E\left[(x_k - x_{k|k})(x_k - x_{k|k})^T\right] \tag{12}$$

Probabilistic data association is applied to accommodate for the possible lack of measurements.

Actions

In this example an action is the selection of the time interval until the next measurement of the target is performed. The action space is a discrete set of possible time intervals:

$$A = \{a = t_r | t_r \in [0.5{:}\,0.5{:}\,5.0]\} \tag{13}$$

This action space could be extended to include waveform selection, such as the number of transmit pulses or the intra-pulse modulation.

State Transition Probability

During the selected time interval, the target is assumed to follow a linear movement with Gaussian noise whereas the platform follows a linear deterministic trajectory. Therefore the state transition equations are:

$$x_k^t = F_{k|k-1}(t_r)x_{k-1}^t + w_{k|k-1}(t_r) \tag{14}$$

$$x_k^p = F_{k|k-1}(t_r)x_{k-1}^p \tag{15}$$

where $F_{k|k-1}$ is the transition matrix, and $w_{k|k-1}$ is a zero-mean white-noise Gaussian distributed variable, with covariance matrix $Q_{k|k-1}$. It can be seen that the action selected influences the transition of the system state between decision stages through the choice of the measurement time interval t_r.

Observations

The radar produces measurements of range r, bearing θ and elevation ϕ, corrupted by Gaussian noise. These are converted into Cartesian coordinates [51] to give the measurement vector $z_k = (\tilde{x}^t, \tilde{y}^t, \tilde{z}^t)^T$.

Observation Likelihood Function

The radar measurements are assumed to be corrupted by Gaussian noise with range, azimuth and elevation standard deviations of σ_r, σ_θ and σ_ϕ respectively. These measurement errors are SNR dependent [52], with a higher SNR leading to a lower standard deviation. The measurement noise in spherical coordinates is then converted into Cartesian coordinates to give the measurement noise covariance $R(x_k)$ [51]. Therefore the observation function is:

$$z_k = H_k\, x_k^t + v_k \tag{16}$$

where H_k is the observation matrix and v_k is a zero-mean white-noise Gaussian distributed variable with covariance matrix $R(x_k)$. The conversion from spherical into Cartesian coordinates is geometry dependent and therefore the Cartesian covariance $R(x_k)$ is dependent on the system state. The observation likelihood function is then:

$$p(z_k|x_k) = N(z_k; H_k x_k, R(x_k)) \tag{17}$$

Reward

In this example, it is desired to minimise the tracking error and track loss while also minimising the resource usage. These conflicting objectives require a trade-off to be found. Consequently, the reward is taken as the tracking utility generated divided by the resource usage. The utility function captures the tracking performance and is defined on the predicted covariance $u: P_{k+1|k} \mapsto U \in [0,1]$:

$$u(P_{k+1|k}) = \begin{cases} 0.0 & \text{if } \sigma\left(P_{k+1|k}\right) \geq 1 \\ 1.0 & \text{if } \sigma\left(P_{k+1|k}\right) \leq 0.2 \\ \left(\dfrac{1-\sigma(P_{k+1|k})}{0.8}\right)^{\eta} & \text{otherwise} \end{cases} \tag{18}$$

where η is a sensitivity parameter and $\sigma(P_{k+1|k})$ calculates the track sharpness [6], [53]. The reward of the belief state b_k is then a function of the utility and the resource:

$$R(b_k, a_k) = \frac{u\left(P_{k+1|k}\right) \cdot t_r}{r_l} \tag{19}$$

where r_l is the resource loading, which is the fraction of radar time used by this task:

$$r_l = \frac{\tau_c}{t_r} \tag{20}$$

with measurement duration τ_c, which is assumed constant for all actions.

3.2.3 Rollout

To solve the POMDP the method of policy rollout is applied, as described in Sec. 2.3.2. The base policy used in this work is to use the same revisit interval as the candidate action for a 5s period in the future, and then to use a 2s revisit interval for the rest of the time horizon, which extends over a total of 25s. The heuristic is chosen on the intuition that the same revisit interval is necessary for a short duration before converging to a regular revisit interval. An example of a rollout execution can be seen in Fig. 3-2 for a 0.5s revisit interval candidate action.

During rollout, expected measurements are generated based on the hypothesis of the system state generated in the rollout branch, and the state dependent measurement covariance. The SNR for the expected measurement is scaled by a beam positioning loss factor that is a function of the track sharpness. This accounts for the inability to correctly direct the radar beam at the target when the track uncertainty is large. The reduced SNR is used to calculate the probability of detection, which is incorporated into the Kalman filter update equation [54].

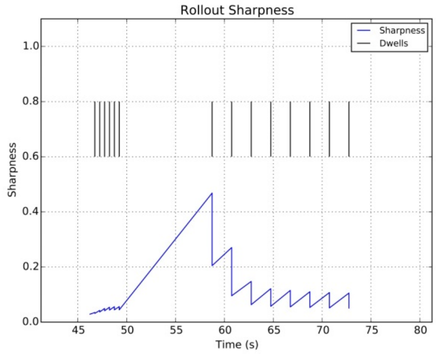

Figure 3-2: Example Rollout Branch.
(Example of a single rollout branch, which represents a hypothesis on the future system evolution. The candidate action is a revisit interval of 0.5s, which is applied for 5s before adopting 2s for the remainder of the time horizon. During the occlusion no measurements are assumed.)

3.2.4 Simulated Results

In the following results, the POMDP with policy rollout described in the previous section is compared against standard adaptive tracking [6], [9], [53], where the track sharpness parameter is set at 0.2. The sensitivity parameter in Eq. 18 is taken as $\eta = 4$.

Fig. 3-3 plots the number of measurements per second that are executed by adaptive tracking and the POMDP for a 2km occlusion. It can be seen that both methods use a high number of measurements at the start of the simulation to initialise the track. It can also be seen that the POMDP anticipates the occlusion by scheduling an increased number of measurements just before the target enters the occluded region. Consequently, the POMDP is able to maintain the tracks during the occlusion and continue tracking once the target is again observable. In contrast, adaptive tracking does not anticipate the occlusion and therefore tracks are lost during the occlusion, which must then undergo a resource expensive track reacquisition when the target is again observable.

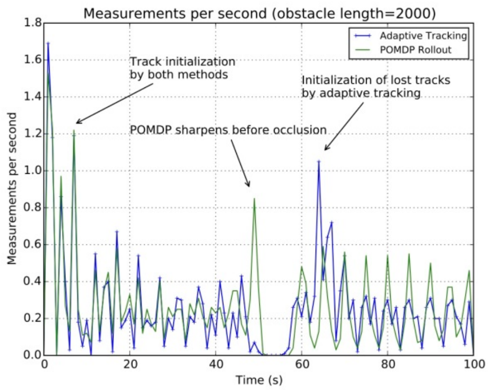

Figure 3-3: Number of measurements per second.
(Number of measurements per second for the POMDP with policy rollout in comparison to adaptive tracking over 100 Monte Carlo runs. The POMDP anticipates the occluded region which leads to a spike in the number of measurements before the occlusion. The adaptive tracking method loses tracks during the occlusion leading to track initialisations when the target is again visible.)

The track sharpness for adaptive tracking and the POMDP method are shown in Fig. 3-4. It can be seen that the POMDP method anticipates the occlusion by sharpening the track before the target enters the unobservable region. In contrast, adaptive tracking does not anticipate the occlusion, leading to much larger track sharpness during the occlusion. The larger track sharpness for adaptive tracking results in track drops and subsequent resource expensive track reacquisitions.

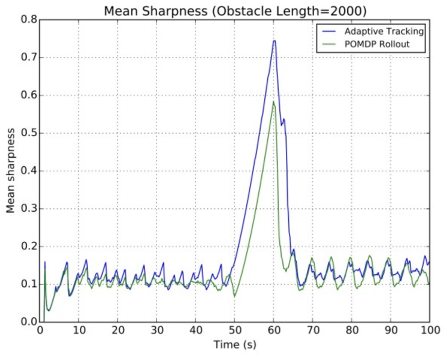

Figure 3-4: Mean Track Sharpness.
(The mean sharpness of both methods, averaged over 100 Monte Carlo runs. It can be seen that the sharpness during the occluded time is lower for the rollout based method, which leads to lower number of track losses.)

In Fig. 3-5 the probability of a track loss, evaluated over 100 Monte Carlo runs, is shown. It can be seen that the probability of a track loss is significantly reduced by the rollout based method, because it anticipates the occlusion and therefore schedules a number of additional measurements shortly before the target is occluded.

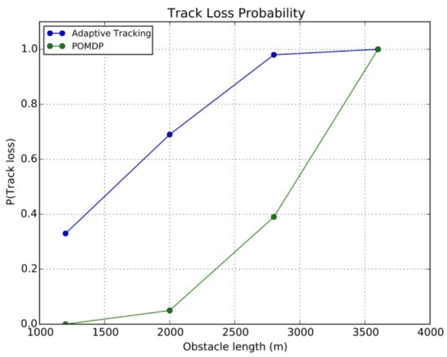

Figure 3-5: Probability of Track Loss.
(The probability of a track loss, based on 100 Monte Carlo runs. It can be seen that the rollout based method has a significantly lower probability of track loss. This is because the rollout anticipates the occlusion, and therefore sharpens the track before the occluded period, leading to a lower number of track losses.)

4.0 SUMMARY

This paper gave an overview of existing techniques that can be said to possess cognitive processes at some level of development. Although some cognitive processes are present in existing methods, there are many cognitive processes that are currently underdeveloped in the context of cognitive radar, such as learning, reasoning, concepts, categorisation, problem solving and language. In order to integrate many cognitive processes a system view can be taken, where the system is structured by information abstraction levels.

This paper then focused on the cognitive process of anticipation, which can be manifested using partially observable Markov decision processes (POMDPs). A POMDP is a problem formulation where actions are selected based on an up to date perception of the system, as well as the expected evolution of the system in the future. Unfortunately, exact solutions to POMDPs are generally intractable, however, many approximate solutions have been developed such as policy rollout. By applying policy rollout to a target tracking control example, it was shown how the process of anticipation could reduce track loss as a target passes through an obscuration.

5.0 REFERENCES

[1] N. Braisby and A. Gellatly, Eds., *Cognitive Psyschology*, 2nd Ed. Oxford University Press, 2012.

[2] M. W. Eysenck and M. T. Keane, *Cognitive Psychology: A Student's Handbook*. Psychology Press, 2015.

[3] F. Barbaresco, J. C. Deltour, G. Desodt, B. Durand, T. Guenais, and C. Labreuche, "Intelligent M3R radar time resources management: Advanced cognition, agility & autonomy capabilities," in *International Radar Conference*, 2009, pp. 1–6.

[4] J. Williams, "Search theory approaches to radar resource allocation," in *Defense Applications of Signal Processing (DASP)*, 2011.

[5] F. Katsilieris, Y. Boers, and H. Driessen, "Optimal search: a practical interpretation of information-driven sensor management," in *5th International Conference on Information Fusion*, 2012, pp. 439 – 446.

[6] G. van Keuk and S. S. Blackman, "On phased-array radar tracking and parameter control," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 1, pp. 186–194, Jan. 1993.

[7] W. D. Blair, G. A. Watson, S. A. Hoffman, and G. L. Gentry, "Benchmark problem for beam pointing control of phased array radar against maneuvering targets," in *American Control Conference*, 1994, vol. 4, pp. 2071–2075.

[8] W. D. Blair, G. A. Watson, T. Kirubarajan, and Y. Bar-Shalom, "Benchmark for radar allocation and tracking in ECM," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 4, pp. 1097–1114, 1998.

[9] W. Koch, "Adaptive parameter control for phased-array tracking," in *SPIE Signal and Data Processing of Small Targets*, 1999, pp. 444–455.

[10] D. J. Kershaw and R. J. Evans, "Optimal waveform selection for tracking systems," *IEEE Transaction on Information Theory*, vol. 40, no. 5, pp. 1536–1550, 1994.

[11] D. J. Kershaw and R. J. Evans, "Waveform selective probabilistic data association," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1180–1188, 1997.

[12] D. Cochran, S. Suvorova, S. Howard, and B. Moran, "Waveform libraries," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 12–21, 2009.

[13] S. P. Sira, Y. Li, A. Papandreou-Suppappola, D. Morrell, D. Cochran, and M. Rangaswamy, "Waveform-agile sensing for tracking," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 53–64, 2009.

[14] S. Haykin, *Cognitive Dynamic Systems: Perception-action Cycle, Radar and Radio*. Cambridge University Press, 2012.

[15] S. Haykin, A. Zia, Y. Xue, and I. Arasaratnam, "Control theoretic approach to tracking radar: First step towards cognition," *Digital Signal Processing*, vol. 21, no. 5, pp. 576–585, 2011.

[16] J. R. Guerci, *Cognitive Radar: The Knowledge-Aided Fully Adaptive Approach*. Artech House, 2010.

[17] J. R. Guerci and E. J. Baranoski, "Knowledge-aided adaptive radar at DARPA: an overview," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 41–50, 2006.

[18] R. Rajkumar, C. Lee, J. Lehoczky, and D. Siewiorek, "A resource allocation model for QoS management," in *18th Real-Time Systems Symposium*, 1997, pp. 298–307.

[19] W. K. Stafford, "MESAR, Sampson & Radar Technology for BMD," in *IEEE Radar Conference*, 2007, pp. 437–442.

[20] W. D. Wirth, *Radar Techniques Using Array Antennas*, 2nd Editio. Institution of Engineering and Technology, 2013.

[21] D. Lambert, "A blueprint for higher-level fusion systems," *Information Fusion*, vol. 10, no. 1, pp. 6–24, 2009.

[22] F. E. White, "A Model for Data Fusion," in *1st National Symposium on Sensor Fusion*, 1988.

[23] A. N. Steinberg, C. L. Bowman, and F. E. White, "Revisions to the JDL data fusion model," in *AeroSense '99*, 1999, pp. 430–441.

[24] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz, and F. White, "Revisiting the JDL Data Fusion Model II," in *7th International Conference on Information Fusion*, 2004.

[25] L. Kester, "Method for Designing Networking Adaptive Interactive Hybrid Systems," *Interactive Collaborative Information Systems*, vol. 281, pp. 401–421, 2010.

[26] F. Smits, A. Huizing, W. van Rossum, and P. Hiemstra, "A cognitive radar network: Architecture and application to multiplatform radar management," in *European Radar Conference*, 2008, pp. 312–315.

[27] E. Chong, C. Kreucher, and A. O. Hero, "Partially observable Markov decision process approximations for adaptive sensing," *Discrete Event Dynamic Systems*, vol. 19, pp. 377–422, 2009.

[28] C. H. Papadimiitriou and J. N. Tsitsiklis, "The Complexity of Markov Decision Processes," *Mathematics of Operations Research*, vol. 12, no. 3, pp. 441 – 450, 1987.

[29] R. D. Smallwood and E. J. Sondik, "The Optimal Control of Partially Observable Markov Decision Processes over a Finite Horizon," *Operations Research*, vol. 21, no. 5, pp. 1071 – 1088, 1973.

[30] J. Pineau, G. Gordon, and S. Thrun, "Point-based value iteration : An anytime algorithm for POMDPs," in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

[31] G. Shani, J. Pineau, and R. Kaplow, "A survey of point-based POMDP solvers," *Autonomous Agents and Multi-Agent Systems*, vol. 27, no. 1, pp. 1–51, Jun. 2012.

[32] J. M. Porta, N. Vlassis, M. T. J. Spaan, and P. Poupart, "Point-Based Value Iteration for Continuous POMDPs," *Journal of Machine Learning Research*, vol. 7, no. Nov, pp. 2329–2367, 2006.

[33] S. Thrun, "Monte Carlo POMDPs," in *Advances in Neural Information Processing Systems 12*, 2000, pp. 1064–1070.

[34] H. Bai, D. Hsu, W. S. Lee, and V. A. Ngo, "Monte Carlo Value Iteration for Continuous-State POMDPs," in *Algorithmic Foundations of Robotics IX*, vol. 68, D. Hsu, V. Isler, J.-C. Latombe, and M. C. Lin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1–16.

[35] H. Bai, D. Hsu, and W. S. Lee, "Integrated perception and planning in the continuous space: A POMDP approach," *The International Journal of Robotics Research*, vol. 33, no. 9, pp. 1288–1302, Jun. 2014.

[36] D. P. Bertsekas and D. a. Castañon, "Rollout algorithms for stochastic scheduling problems," *Journal of Heuristics*, vol. 5, no. 1, pp. 89–108, 1999.

[37] S. Ross, J. Pineau, S. Paquet, and B. Chaib-draa, "Online Planning Algorithms for POMDPs," *Journal of Artificial Intelligence Research*, vol. 32, pp. 663–704, 2008.

[38] E. K. P. Chong, "POMDP Approximations using Simulation and Heuristics," in *Foundations and Applications of Sensor Management*, 2008, pp. 95–119.

[39] H. S. Chang, R. Givan, and E. K. P. Chong, "Parallel rollout for online solution of partially observable Markov decision processes," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 14, no. 3, pp. 309–341, 2004.

[40] D. A. Castanon and L. Carlin, "Stochastic Control Theory for Sensor Management," in *Foundations and Applications of Sensor Management*, 2008, pp. 7–32.

[41] C. Kreucher, K. Kastella, and A. O. Hero, "Sensor management using an active sensing approach," *Signal Processing*, vol. 85, no. 3, pp. 607–624, 2005.

[42] J. Wintenby, "Resource allocation in airborne surveillance radar," Chalmers University of Technology, 2003.

[43] J. Wintenby and V. Krishnamurthy, "Hierarchical resource management in adaptive airborne surveillance radars," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 2, pp. 401–420, 2006.

[44] Y. Li, L. W. Krakow, E. K. P. Chong, and K. N. Groom, "Approximate stochastic dynamic programming for sensor scheduling to track multiple targets," *Digital Signal Processing*, vol. 19, no. 6, pp. 978–989, Dec. 2009.

[45] S. A. Miller, Z. A. Harris, and E. K. Chong, "A POMDP Framework for Coordinated Guidance of Autonomous UAVs for Multitarget Tracking," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, p. 724597, Jan. 2009.

[46] S. Ragi and E. Chong, "UAV path planning in a dynamic environment via partially observable Markov decision process," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 4, pp. 2397–2412, 2013.

[47] B. F. La Scala, W. Moran, and R. J. Evans, “Optimal adaptive waveform selection for target detection,” in *International Radar Conference*, 2003, pp. 492–496.

[48] B. F. La Scala and B. Moran, “Optimal target tracking with restless bandits,” *SPIE Signal Processing, Sensor Fusion, and Target Recognition XV*, vol. 16, no. 5, pp. 479–487, Sep. 2006.

[49] B. L. Scala, M. Rezaeian, and B. Moran, “Optimal adaptive waveform selection for target tracking,” in *8th International Conference on Information Fusion*, 2005, vol. 1.

[50] A. B. Charlish and F. Hoffmann, “Anticipation in Cognitive Radar using Stochastic Control,” in *IEEE International Radar Conference*, 2015.

[51] F. K. Fletcher and D. J. Kershaw, “Performance Analysis of Unbiased and Classical Measurement Conversion Techniques,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 4, pp. 1441–1444, 2002.

[52] D. K. Barton, *Radar Systems Analysis and Modelling*, Rev. 1st. Artech House, 2004.

[53] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.

[54] T. Fortmann, Y. Bar-Shalom, M. Scheffe, and S. Gelfand, “Detection thresholds for tracking in clutter--A connection between estimation and signal processing,” *IEEE Transactions on Automatic Control*, vol. 30, no. 3, pp. 221–229, 1985.