Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

Look versus Leap: Computing Value of Information with High-Dimensional Streaming Evidence

Stephanie Rosenthal*, **Dan Bohus, Ece Kamar, Eric Horvitz**
Microsoft Research, Redmond, WA USA
srosenth@cs.cmu.edu, {dbohus, eckamar, horvitz}@microsoft.com

Abstract

A key decision facing autonomous systems with access to streams of sensory data is whether to act based on current evidence or to wait for additional information that might enhance the utility of taking an action. Computing the value of information is particularly difficult with streaming high-dimensional sensory evidence. We describe a *belief projection* approach to reasoning about information value in these settings, using models for inferring future beliefs over states given streaming evidence. These belief projection models can be learned from data or constructed via direct assessment of parameters and they fit naturally in modular, hierarchical state inference architectures. We describe principles of using belief projection and present results drawn from an implementation of the methodology within a conversational system.

1 Introduction

Real-time adaptive systems, such as self-driving cars, monitoring systems, and conversational agents, reason continuously under uncertainty: they probabilistically estimate the state of the world based on data collected via multiple sensors and use these state estimates to guide action selection. Sensors and inferences can be characterized by different levels of accuracy and by stochastic delays, and collecting additional evidence can lead to improved state estimates at the cost of delayed action. For instance, face identification often becomes more accurate as a person approaches a camera. Significant tradeoffs can arise between the decision to act immediately based on the available evidence or to wait for additional evidence to accumulate for the future action. To resolve such tradeoffs, a system must reason about the expected benefits and costs of future courses of actions, including gathering additional information under uncertainty.

Exact computation of the value of information with streaming sensory data is challenging for several reasons. Real-time systems often harness high-dimensional streams of evidence, which introduce challenges in training accurate generative observation models. As an example, an embodied conversational agent can leverage pixels in RGB video streams, depth maps, and audio signals to track the location, intentions and goals of people in the surrounding environment. Building generative models that accurately predict future pixels, depth maps, and audio signals based on the previous series of such rich observations is a challenging task. In fact physically situated conversational systems often rely on hierarchies of conditional models that abstract the high-dimensional streaming sensory evidence into fewer, lower-dimensional *percepts*. For instance, a conversational agent may use vision models to track and identify people in its vicinity based on an analysis of pixels in a video stream. These trajectory and identity percepts can be used to infer peoples' goals and intentions, and ultimately to drive decisions about interaction. Even when relying on the abstraction of high-dimensional sensors, performing belief updates by iteratively combining predictions of generative observation models can lead to an accumulation of errors over long periods of time.

We address the challenge of constructing information-gathering policies for systems that rely on high-dimensional streaming sensor data. As we will discuss in more detail, traditional approaches for computing the value of information do not perform well in these settings as they use generative models for predicting future observations that are hard to train and reason with. We propose an approach we refer to as *belief projection*, in which direct conditional models are used to predict future beliefs from a prior stream of observations. We show how these models can be used to resolve the tradeoffs between acting immediately, waiting for more evidence to accumulate, or, more generally, orchestrating which sensors should be activated at a given time. The proposed belief projection models can be learned automatically, via self-supervision, and fit naturally into hierarchical inferential architectures. We demonstrate a practical implementation of these ideas in a physically situated embodied conversational agent that uses natural language to interact with people. The proposed approach enables a mixed-initiative engagement policy in this system, and opens new opportunities for solving other real-world decision problems.

2 Related Work

Tradeoffs between acting and waiting can be resolved by computing the expected value of information (VOI). VOI has

*Stephanie Rosenthal contributed to this research during an internship at Microsoft Research.

been also used in prior work for evaluating the value for performing additional computation [Horvitz, 1987] and for evaluating the value of precomputing answers to future problem instances with currently available resources [Horvitz, 2001].

In settings with large state spaces and high-dimensional streaming evidence, exact computation of VOI is challenging for several reasons. First, the computation requires reasoning about an exponentially growing tree of future evidence gathering actions and outcomes [Heckerman *et al.*, 1993], which quickly renders the problem intractable. To address this issue, various approximations of VOI have been proposed. Most VOI approximations employ a greedy analysis, assuming that only a single piece of evidence will be observed [Ben-Bassat, 1978; Heckerman *et al.*, 1992] before action is taken. If the problem has a submodular structure, greedy VOI approaches have been shown to have well-defined error bounds [Krause and Guestrin, 2007]. Non-myopic VOI analyses have been formulated [Heckerman *et al.*, 1993; Bilgic and Getoor, 2011; Liao and Ji, 2008] for situations where sequential observations are independent. When future evidence can perfectly predict the state, approximate algorithms [Armstrong-Crews and Veloso, 2008] can be used. Algorithms have also been proposed for problems with decomposable structure [Boutilier *et al.*, 1999], and for problems with a search space that grows linearly in the horizon [Hajishirzi *et al.*, 2009]. Decisions about when to terminate evidence collection have also been studied in optimal stopping problems [Peškir and Širjaev, 2006]. Tractable solutions developed for this class of problems are not applicable to our setting as the aggregation of evidence can have arbitrary influences on beliefs.

The challenge of acting under uncertainty with streaming evidence has been noted as a key AI research problem [Selman *et al.*, 1996]. Previous work on computing VOI in such settings has focused on myopic analyses and domain-specific heuristics [Oliver and Horvitz, 2003]. We propose a new approach to reasoning about information value amidst streams of high-dimensional evidence. The proposed approach is myopic but includes actions for waiting different amounts of time for new streaming evidence.

Besides the exponentially growing search tree, a second important challenge in computing VOI with high-dimensional evidence has to do with the use of generative observation models [Kaelbling *et al.*, 1998]. Constructing such models is difficult for domains with high-dimensional observation spaces. In addition, it has been shown [Kamar *et al.*, 2012] that predicting future beliefs by iteratively applying generative observation models leads to accumulation of errors in belief predictions. Predictive state representations have been proposed as an alternative to generative approaches [Littman *et al.*, 2002]. However, these representations are also prone to error accumulation as a result of iterative updating. We introduce belief projection, where direct conditional models are used for tractable and accurate predictions of future beliefs.

3 Approach

To resolve the tradeoff between acting immediately and waiting to collect additional observations, systems need to consider the uncertainty of their current state and of future observations. The tradeoff can be formalized using a POMDP, or equivalently a belief MDP [Kaelbling *et al.*, 1998]. $s_t \in S$ represents the state of the world at time t. Ψ_t is a sequence of evidence vectors that are collected by the system from time 1 to time t, $\Psi_t = \{\psi_i\}_{i=1:t}$. $b_t(s_t)$ represents the system's belief at time t about being in state s_t. $p(b_t|\Psi_t)$ is the probability of b_t given evidence collected. The set of actions A is composed of a set of domain actions A_d and the action a_w which represents waiting a single time step to collect additional information. The system transitions to a terminal state after taking a domain action. $p(\Psi_{t+1}|\Psi_t)$ is the probability of future evidence after taking action a_w. $R(s_t, a)$ is the system's reward for taking action a at state s_t. The expected reward for belief state b_t is computed $r(b_t, a) = \sum_{s_t} b_t(s_t) \cdot R(s_t, a)$.

The VOI at any belief state b_t is the difference between V_{wait}, the value for waiting, and V_{act}, the value of acting. V_{wait} and V_{act} are computed with the Bellman equation:

$$V_{act}(b_t) = \max_{a \in A_d} r(b_t, a) \tag{1}$$

$$V_{wait}(b_t) = r(b_t, a_w) + \sum_{\Psi_{t+1}} p(\Psi_{t+1}|\Psi_t) \cdot \sum_{b_{t+1}} p(b_{t+1}|\Psi_{t+1}) \cdot V^*(b_{t+1}) \tag{2}$$

where $V^*(b_t) = \max(V_{act}(b_t), V_{wait}(b_t))$. A positive VOI indicates that it is beneficial to wait for more evidence.

The number of belief states that need to be explored for the exact computation of VOI grows exponentially with the depth of lookahead, rendering exact computation intractable, especially when evidence arrives incrementally in a streaming manner. To overcome this complexity, we introduce an approximation procedure that converts the sequential decision problem into a myopic problem, while still considering the utility of taking sequences of information gathering actions. We move beyond the recursive consideration of sequences and introduce longer duration wait actions, $a_{w(k)}$, which represent waiting for k time steps and then taking a single domain action and terminating. The reward for taking these actions can be approximated from existing reward functions, *e.g,* $r(b_t, a_{w(k)}) = k \cdot r(b_t, a_w)$, or can be elicited directly from domain experts. The value of taking $a_{w(k)}$ is calculated as follows, under the assumption that the system will take the best domain action at the end of waiting:

$$V_{wait(k)}(b_t) = r(b_t, a_{w(k)}) + \sum_{\Psi_{t+k}} p(\Psi_{t+k}|\Psi_t) \cdot \sum_{b_{t+k}} p(b_{t+k}|\Psi_{t+k}) \cdot V_{act}(b_{t+k}) \tag{3}$$

The value of waiting ($V_{wait}(b_t)$) is approximated as the maximum of values associated with the different wait durations.

Although greedy, this computation is still intractable as it requires a model that predicts future sensory evidence, $p(\Psi_{t+k}|\Psi_t)$. Building a generative model that predicts future observations is intractable because of the high-dimensionality and streaming nature of the sensory evidence ψ. Complex systems that harness perceptual machinery, such as cameras and microphones, use off-the-shelf non-decomposable components for making state inferences $p(s_t|\Psi_t)$, where ψ_i is high-dimensional, *e.g.,* pixels for face tracking or raw audio for speech recognition. Alternative formulations using

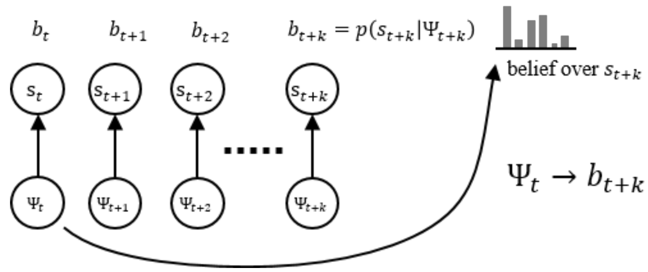

Figure 1: Training setup for belief projection models.

factorizations of $p(\Psi_{t+k}|\Psi_t)$ into $p(\Psi_{t+k}|s_{t+k}) \cdot p(s_{t+k}|\Psi_t)$ encounter similar tractability challenges.

To address this computational challenge, we introduce belief projection. Belief projection models are direct conditional models of the form $p(b_{t+k}|\Psi_t)$ that predict the future belief at time $t + k$ based on the evidence Ψ_t collected up to time t. These models do not require predictions of future sensory evidence streams, and can be used to estimate the value of wait actions:

$$V_{wait(k)}(b_t) = r(b_t, a_{w(k)}) + \sum_{b_{t+k}} p(b_{t+k}|\Psi_t) \cdot V_{act}(b_{t+k}) \quad (4)$$

The belief projection model $p(b_{t+k}|\Psi_t)$ can be trained in a supervised manner without requiring manual labeling (see Figure 1). The training data, consisting of pairs of the form $\Psi_t \rightarrow b_{t+k}$, can be collected automatically at runtime by recording Ψ_t, the sensory evidence collected up to time t, and the output of the state inference model at time $t + k$ (see Figure 1). For each data point (Ψ_t, b_{t+k}), the features Ψ_t describe the sensory evidence collected up to time t. The corresponding label, b_{t+k}, consists of the output of the state inference models at some future time $t + k$; the training label is a belief over the state s_{t+k}. For instance, if the state is binary, *i.e.,* $s_t \in \{0, 1\}$, the belief over s_t is defined over the unit 1 simplex, *i.e.,* $b_t \in \Delta^1$, which is the $[0, 1]$ real interval. In this case, the belief projection model constructs a probability distribution over this simplex, or over the [0, 1] interval.

A possible approach to constructing belief projection models is to employ a mixture of Beta distributions (see the application described in Section 4). The model parameters can be trained from data in a maximum likelihood manner. An alternative is to discretize the $[0, 1]$ interval into several bins, treat the problem as multinomial classification, and to build a model via discriminative learning techniques such as maximum entropy or decision trees. The complexity of the learning problem increases as the size of the original state space increases. For instance, if instead of being binary, the state is a multinomial variable with d possible values, the belief over s_t is defined over the unit $d - 1$ simplex, *i.e.,* $b_t \in \Delta^{(d-1)}$. The state projection inference model could be constructed in this case as a mixture of Dirichlet distributions.

Finally, we note that the proposed approach requires that we sum over all possible beliefs b_{t+k}. In practice, a tractable solution for computing this sum (or integral) is required. One approach that works well when the underlying state space is small is to discretize the belief space (the *simplex*) into a number of bins, and to sum over the corresponding probabilities. Another alternative is to construct belief projection models with parametric forms that allows for analytic integration. We have explored the use of non-parametric techniques (decision trees and random forests) for learning belief projection models that can directly provide belief samples which can be used in the VOI computation.

As we mentioned earlier, belief projection is well-suited for complex systems constructed via the coupling of multiple, modular inference components into hierarchical state inference architectures. In such systems, lower-level components, such as speech recognition and face tracking and identification, are usually trained and optimized separately, prior to integration in a larger application. These components abstract the high-dimensional streaming sensory evidence, such as raw audio and video data, into fewer lower-dimensional percepts, such as words spoken and the location and identity of a person. The outputs of these perceptual models are then used as inputs for higher-level domain-specific inferences models that reason about goals, activities, and other relevant state variables, which ultimately drive interaction decisions. An example is shown in Figure 3 and discussed in more detail in the next section. Belief projection is a natural fit for computing VOI for such modular architectures as models can be trained independently for each low-level perceptual inference model. The projected beliefs over the percepts can then be used to infer the corresponding beliefs of the high-level state.

4 Application

We now review an implementation of the methodology described earlier to support a mixed-initiative engagement policy in a deployed system that has been serving as an automated secretary for nearly two years.

4.1 The Assistant

The Assistant is a multimodal interactive kiosk displaying an animated avatar head. The system is stationed outside the office of an employee at our organization (see Figure 2 part (a)) and has been serving as an automated secretary. The system has access to the owner's calendar, computer activity, and the wifi fingerprints of devices on the owner's network, and continuously makes probabilistic forecasts about the owner's availability and arrival for meetings via subsystems developed as separate research projects [Horvitz *et al.*, 2002; 2004]. The Assistant can interact via spoken language with visitors who stop by the owner's office and handles a variety of administrative tasks, such as providing information about the activities, whereabouts, and future availability of its owner, scheduling meetings, and relaying messages. The system leverages an array of sensors and makes real-time inferences about people in its proximity, including their identities, activities and goals, in support of natural and effective interactions with visitors.

4.2 Engagement Problem

An important challenge for the Assistant is managing its engagement with people. *Engagement* is the process by which participants in a conversation coordinate their actions to initiate, maintain and terminate their interactions [Sidner *et al.*,

(a) (b) (c) (d)

Figure 2: (a) The Assistant outside the owners office. (b) Visitor walks through the Assistant's field of vision. (c) Visitor heads towards the chair to wait for owner. (d) Visitor sits down in Assistant's field of view.

2005; Bohus and Horvitz, 2009]. Prior versions of the Assistant used a conservative, heuristic engagement policy: the system waited for users to initiate engagement by entering in an *f-formation* [Kendon, 1990] with it, *i.e.,* by approaching and standing closely in front of it. This policy was designed to minimize false positive cases where the Assistant would accidentally initiate engagement with someone walking by or standing nearby but talking to someone in another office. At the same time, we had noticed that people who are waiting for the owner (to return to his office, or to become available) often bypass the system and sit in nearby chairs (see Figure 2 (a)) or talk to others while waiting. In these situations, the Assistant's engagement policy did not allow it to proactively initiate engagement, which led to missed opportunities to engage a waiting person in dialog. Such missed opportunities can be costly. For example, the system may know that the owner is running late for a scheduled meeting, and would like to relay this information if the person waiting is scheduled to meet with the owner. Without knowledge of the delay, the visitor may leave in frustration.

The engagement problem is a sequential decision-making problem with streaming evidence. The Assistant can partially observe the state of the world: it cannot perfectly observe a visitor's goals or intentions to engage with the system. However, the Assistant has access to streaming observations, which provide additional evidence about the true state of the world over time. The observations are complex, composed of information from multiple sensors including a camera and a microphone, and the domain of observations can be very large. The decision to reach out and engage someone near the system depends on inferences about the state of the world based on the streaming and high-dimensional observations.

At each time step, the Assistant makes a decision between acting to engage right away, and waiting to gather additional sensory evidence in support of better decisions. In addition, the Assistant can take actions to seek additional evidence in real time. Specifically, the system can take and send a snapshot of the visual scene to human volunteers and ask them to help identify the person in the scene. When deciding how to act, the Assistant needs to reason about multiple sources of uncertainty: uncertainty about the streaming observations it may collect, about when the responses to information gathering actions will arrive, and about whether and when the person may leave the scene (and thus curtail an opportunity to engage in conversation).

The methodology described earlier enables a mixed-initiative engagement policy which allows the Assistant to initiate engagement proactively with users (even at a distance) when it is confident about the state of the user, and to wait or trigger external queries for additional evidence when it is uncertain. This engagement policy hinges on inferences about the current state and future observations, and trades off the value of additional sensory evidence that is likely to be accumulated in the future with the costs of engaging people that are not looking for the owner, and the costs of missed opportunities to engage visitors before they leave.

4.3 Formalization

The engagement problem described above can be formalized in terms of a belief MDP. The state S relevant for making engagement decisions subsumes three variables: (1) the users *Presence* (P), which can take two possible values: *present*, denoting that the user is still present in the scene, or *not-present*, denoting that the user has left the scene; (2) the *User's Engagement Action* (UEA), which can take two possible values: *engaging*, denoting that the user is entering in an f-formation with the Assistant, or *not-engaging* otherwise; and (3) the users *Goal* (G), which can take two possible values: *looking-for-owner*, denoting that the user is looking for the owner, or *other* if this is not the case. The Assistant can observe whether a user is present, but cannot fully observe the user's engagement action or goal. The Assistant keeps a belief state representing its belief about the non-observable variables of the state. The belief is inferred based on a hierarchical graphical model shown in Figure 3, which leverages three lower-level percepts: *F-Formation* (FF), which indicates whether or not a user is entering in an f-formation with the Assistant; *Activity* (A), which indicates whether or not the user is approaching the Assistant; and *On-Calendar* (OC), which indicates whether the user has a meeting with the owner that has started or is about to start. These lower-level percepts are in turn inferred based on direct conditional models that leverage even lower-level high-dimensional observations obtained from sensors such as sustained attention, the trajectory of the face, speed of movement, proximity, centrality of location, face identification information, as well as the owner's calendar information.

The engagement problem includes two domain actions: *Engage*, in which the Assistant engages the user immediately, and *DontEngage*, in which the Assistant decides to not en-

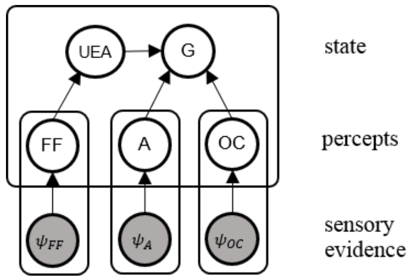

Figure 3: Hierarchical graphical model for state inference.

States S		System Action (A)	Utility
Engagement Action (UEA)	Goal (G)		
engaging	⟨any⟩	Engage	1.00
		DontEngage	0.05
not-engaging	looking-for-owner	Engage	0.75
		DontEngage	0.25
	other	Engage	0.10
		DontEngage	1.00

Table 1: Utilities for state and action combinations.

gage the user. If the user is not present, the Assistant can only take the *DontEngage* action. The decision-making problem terminates after taking any of the domain actions. Evidence collection actions include: *Wait(t)* to collect additional sensory information, and *AskAndWait(t)* to ask an external source whether the user is the person on the owner's calendar and also collect sensory information while waiting for the response, where t ranges from 1 to 100 seconds.

The reward function represents the utility for taking an action at a given state. The rewards for taking *Engage* and *DontEngage* actions at different states elicited from the Assistant's owner are shown in Table 1. The current policy in the Assistant considers a 0.05 cost for taking action *AskAndWait(t)*.

In belief MDPs, the transition model represents the probability of transitioning to a belief state after taking an action. Traditional methods for predicting future belief states would use observation models for predicting future evidence. Deliberating amidst high-dimensional streaming evidence makes these methods infeasible. Instead, we combine predictions of multiple belief projection models, as outlined in Section 3, to predict future belief states. Three belief projection models predict the future values of FF, A and OC variables. The hierarchical model given in Figure 3 is used to turn beliefs over these perceptual variables to beliefs over binary UEA and G variables. The belief over each of these binary variables is a 1-dimensional simplex, *i.e.,* the interval $[0, 1]$. In this case, we construct the three belief projection models heuristically, based on mixtures of Beta distributions; we present these models in Subsection 4.5, where we illustrate their implementation and function with a concrete example.

To determine the transition probabilities, the predictions of belief projection models are combined with predictions of a model for the likelihood that a user may leave the scene. This inference is based on the time since the actor was detected, via a mixture of two linear hazard rate distributions: the first component has a mean of ~ 4 seconds and models people that simply pass through the corridor and the second component has a mean of ~ 300 seconds and models people that sit in an alcove near the Assistant.

When the *AskAndWait(t)* action is taken, the transition model takes into account the likelihood that the response will arrive sometime in the future and the likelihood about the content of the message if it arrives. The first likelihood is modeled via a log-normal distribution with a mean of 40 seconds. The likelihood of the content of the message is modeled with the Assistant's current prediction of the OC variable. Once the message arrives, the Assistant updates its belief about the OC variable with the content of the message.

4.4 Selecting Actions

Based on the methodology outlined in Section 3, the Assistant computes the expected utilities for taking any of the domain actions, and for taking any of the information-gathering actions followed by the best domain action. The expected utilities for domain actions reflect the Assistant's belief about immediate rewards. The expected utilities for observation gathering actions combine the potential benefits of improved estimates of the state of the world with the potential costs of external queries or for missing the opportunity to engage if the user leaves the scene.

The Assistant recomputes the utilities at every time step and chooses the action with the highest expected utility. With this replanning approach, the Assistant may choose a particular action like *Wait(10)* at a certain time, while at the next time step the action selected might change based on the accumulated evidence, *e.g.,* to something like *Engage* or *Wait(50)*.

4.5 Sample Trace

We now illustrate the methods with a sample trace of the Assistant's operation. In the example, a visitor approaches the office where the Assistant is stationed (see Figure 2(b)), passes by the system (see Figure 2(c)) and sits down in a nearby chair (see Figure 2(a,d)).

Between times t_1 and t_2, as the visitor approaches, the width of the face (Figure 4A), as well as the probability of f-formation and approaching (Figure 4D) are increasing; the Assistant is uncertain about whether this visitor is on-calendar (Figure 4D). Based on the lower-level perceptual

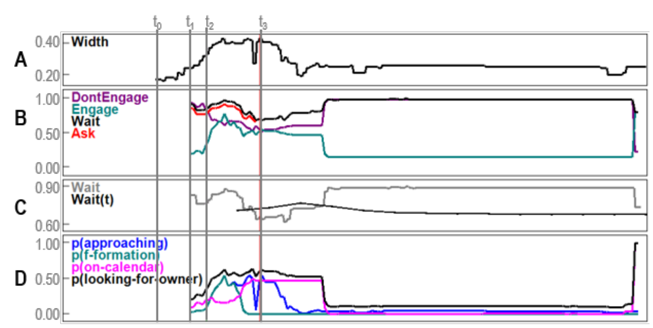

Figure 4: Sample traces: face width (A), action utilities (B, C), and inferences (D).

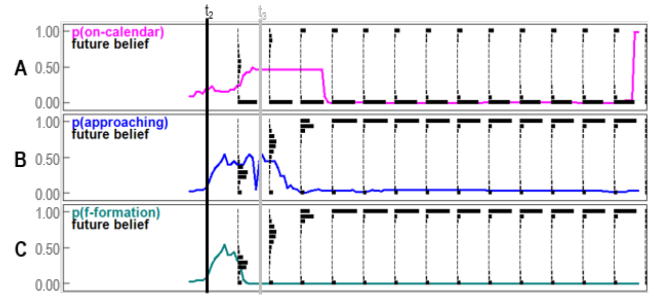

Figure 5: Projected percept beliefs computed for t_2 for *On-Calendar* (A), *Activity* B, *F-Formation* (C), and *Wait(t)* computed at t_2.

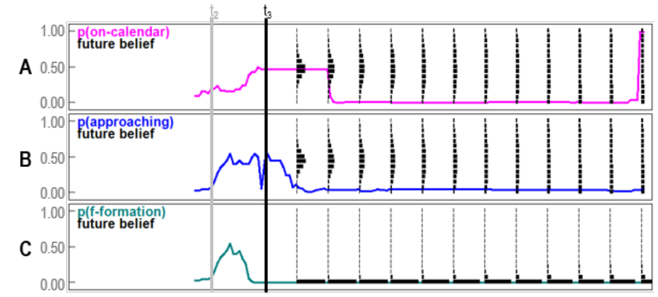

Figure 6: Projected percept beliefs computed for t_3 for *On-Calendar* (A), *Activity* B, *F-Formation* (C).

evidence, the state inference indicates that at time t_2 the marginal probability that the visitor is *looking-for-owner* is 0.33 (Figure 4D). As the visitor is approaching, the expected utility for *Engage* has been steadily increasing and for *DontEngage* has been decreasing (Figure 4B).

The Assistant also computes the utility of *Wait(t)* and *AskAndWait(t)*. For each time t, these computations are based on projection models for the *On-Calendar*, *Activity* and *F-Formation* percepts. Figures 5A, B, and C show the projected future beliefs (black histograms at every future time point), as computed at time t_2. The actual beliefs (constructed at the future points) are also shown as a solid, colored line.

The projected beliefs for the *On-Calendar* percept, shown as black histograms in Figure 5A, are constructed based on a mixture of two Beta distributions. The mixture model is formulated such that, if the actor is getting closer the system (as in our example), the *On-Calendar* perceptual inference is more likely to output future beliefs with more certainty, concentrated towards 0 and 1. The projected beliefs for the *Activity* and *F-Formation* percepts are computed similarly and displayed in Figures 5B and C. These figures indicate that, if the Assistant waits, there will be reduced uncertainty over whether the person is on the calendar, whether they are approaching, and whether they are entering in an f-formation. The computation for the expected utility of *Wait(t)* performed at time t_2 integrates over these predicted future beliefs, and also takes into account the probability that the actor will disappear. Figure 4C shows the resulting expected utility of *Wait(t)* for different values of t. The maximum expected utility is attained for a wait time of t=3 seconds, and corresponds to the value for the *Wait* action shown in gray in Figure 4C. Similarly, the computation for the expected utility of *AskAndWait(t)* integrates over the predicted future beliefs, as well as over the probability that the response might arrive by time t, and takes into account the likelihood of different responses.

As Figure 4B shows, while initially the expected utility on *Engage* increases, and even exceeds the expected utility of *DontEngage* shortly thereafter, the expected utility of *Wait* is even larger; the system infers that waiting is most beneficial since the person is getting closer and uncertainties about their identity, and, ultimately their goals, will likely be resolved and a better decision can be made in the near future.

Next, the visitor passes by the Assistant and sits in a nearby chair (see Figure 2(d)). In Figure 6A, B, and C, we show the projected beliefs for the *On-Calendar*, *Activity* and *F-Formation* percepts, as computed at time t_3. As the person is no longer moving closer, the projected beliefs indicate that there is not much to gain by waiting. At t_3, the expected utility of *AskAndWait* exceeds the expected utility of *Wait* (see Figure 4B), as the perceptual belief projection model indicates that the identity will not be known better in the future and the visitor will not likely leave immediately. The Assistant launches an external query about the visitor's identity.

From this point forward, the *AskAndWait* action is no longer evaluated, but the utility computation for the *Wait(t)* action also reflects the fact that the response to the information gathering action might arrive. When this happens a few seconds later, at time t_4, the system finds that the visitor is indeed the person expected and the corresponding probability for *on-calendar* increases to 1.0 (see Figure 4D). The maximum expected utility action becomes *Engage* (see Figure 4B), and the system proactively engages the visitor that is at this point still waiting in the chair (see Figure 2D), by saying "Pardon me, are you looking for [Owner's Name]?"

5 Conclusions and Future Work

We introduced belief projection for computing the value of information in systems that operate with high-dimensional streaming sensory evidence. The approach relies on developing direct conditional models to predict future beliefs based on current evidence. Belief projection can be used to resolve tradeoffs between acting immediately and waiting for more evidence to accumulate. The methods are well-suited for systems that use hierarchical architectures for making state inferences. We implemented belief projection in a deployed interactive agent and illustrated how the methodology enables mixed-initiative engagement policies. Using belief projection models constructed with heuristic parameters, the system is able to deliberate about the value of waiting for more information from sensors, soliciting help in real time from experts, or acting immediately.

We are currently exploring the learning of the projection models from case libraries of sensory data via non-parametric methods. Another direction for future work is investigating whether belief projection can be applied to broader decision-making problems with arbitrary sequences of actions.

References

[Armstrong-Crews and Veloso, 2008] N. Armstrong-Crews and M. Veloso. An approximate solution for oracular pomdps. In *ICRA '08*, pages 3346 – 3352, 2008.

[Ben-Bassat, 1978] M. Ben-Bassat. Myopic policies in sequential classification. *IEEE Trans. Comput.*, 27:170–178, 1978.

[Bilgic and Getoor, 2011] M. Bilgic and L. Getoor. Value of information lattice: Exploiting probabilistic independence for effective feature subset acquisition. *Journal of Artificial Intelligence Research*, 41:69–95, 2011.

[Bohus and Horvitz, 2009] D. Bohus and E. Horvitz. Models for multiparty engagement in open-world dialog. In *Meeting of the Special Interest Group on Discourse and Dialogue*, pages 225–234, 2009.

[Boutilier *et al.*, 1999] C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94, 1999.

[Hajishirzi *et al.*, 2009] H. Hajishirzi, A. Shirazi, J. Choi, and E. Amir. Greedy algorithms for sequential sensing decisions. In *International Joint Conference on Artificial Intelligence*, pages 1908–1915, 2009.

[Heckerman *et al.*, 1992] D. Heckerman, E. Horvitz, and B. Nathwani. Toward normative expert systems: The pathfinder project. Technical report, Knowledge Systems Laboratory, Medical Computer Science, Stanford University, 1992.

[Heckerman *et al.*, 1993] D. Heckerman, E. Horvitz, and B. Middleton. An approximate nonmyopic computation for value of information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 292–298, 1993.

[Horvitz *et al.*, 2002] E. Horvitz, P. Koch, C. M. Kadie, and A. Jacobs. Coordinate: Probabilistic forecasting of presence and availability. In *Uncertainty and Artificial Intelligence*, pages 224–233, 2002.

[Horvitz *et al.*, 2004] E. Horvitz, J. Apacible, and P. Koch. Busybody: Creating and fielding personalized models of the cost of interruption. In *Conference on Computer Supported Cooperative Work*, pages 507–510, 2004.

[Horvitz, 1987] E. Horvitz. Reasoning about beliefs and actions under computational resource constraints. In *3rd Conference on Uncertainty in Artificial Intelligence*, pages 429–444, 1987.

[Horvitz, 2001] E. Horvitz. Principles and applications of continual computation. *Artificial Intelligence*, 126:159–196, 2001.

[Kaelbling *et al.*, 1998] L. Kaelbling, M. Littman, and A Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.

[Kamar *et al.*, 2012] E. Kamar, S. Hacker, , and E. Horvitz. Combining human and machine intelligence in large-scale crowdsourcing. In *AAMAS*, pages 467–474, 2012.

[Kendon, 1990] A. Kendon. *Conducting Interaction: Patterns of behavior in focused encounters*. Cambridge University Press, 1990.

[Krause and Guestrin, 2007] A. Krause and C. Guestrin. Near-optimal observation selection using submodular functions. *National Conference on Artificial Intelligence*, 22(2):1650–1654, 2007.

[Liao and Ji, 2008] W. Liao and Q. Ji. Efficient non-myopic value-of-information computation for influence diagrams. *International Journal of Approximate Reasoning*, 49:436–450, 2008.

[Littman *et al.*, 2002] Michael L Littman, Richard S Sutton, and Satinder Singh. Predictive representations of state. *Advances in neural information processing systems (NIPS)*, 14:1393–1400, 2002.

[Oliver and Horvitz, 2003] N. Oliver and E. Horvitz. Selective perception policies for guiding sensing and computation in multimodal systems: A comparative analysis. In *International Conference on Multimodal Interfaces*, 2003.

[Peškir and Širjaev, 2006] Goran Peškir and Al'bert Nikolaevič Širjaev. *Optimal stopping and free-boundary problems*. Springer, 2006.

[Selman *et al.*, 1996] B. Selman, R. A. Brooks, T. Dean, E. Horvitz, T. M. Mitchell, and N. J. Nilsson. Challenge problems for artificial intelligence. In *AAAI-96*, pages 1340–1345, 1996.

[Sidner *et al.*, 2005] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, , and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164, 2005.