EPOC: Efficient Perception via Optimal Communication

Masoumeh Heidari Kapourchali, Bonny Banerjee
Institute for Intelligent Systems, and Department of Electrical and Computer Engineering, University of Memphis
Memphis, TN 38152, USA
{mhdrkprc, bbnerjee}@memphis.edu

Abstract

We propose an agent model capable of actively and selectively communicating with other agents to predict its environmental state efficiently. Selecting whom to communicate with is a challenge when the internal model of other agents is unobservable. Our agent learns a communication policy as a mapping from its belief state to *with whom to communicate* in an online and unsupervised manner, without any reinforcement. Human activity recognition from multimodal, multisource and heterogeneous sensor data is used as a testbed to evaluate the proposed model where each sensor is assumed to be monitored by an agent. The recognition accuracy on benchmark datasets is comparable to the state-of-the-art even though our model uses significantly fewer parameters and infers the state in a localized manner. The learned policy reduces number of communications. The agent is tolerant to communication failures and can recognize unreliable agents through their communication messages. To the best of our knowledge, this is the first work on learning communication policies by an agent for predicting its environmental state.

I. Introduction

This paper investigates how *an* agent can optimally use other agents for predicting the state of its environment. The assumption is that, interacting agents might have distinct goals but can still benefit from each other's knowledge. We propose an agent model that learns to communicate selectively with other agents to predict its environmental state.[1]

We model communication as active perception (Bajcsy, Aloimonos, and Tsotsos 2018). This allows an agent to actively and selectively sample (or *communicate with*) other agents. Communication makes causal knowledge acquistion efficient by allowing to: (1) share causal knowledge regarding the same event even though the observations are from different sensors in space, time or modality, and (2) acquire high-level causal knowledge directly from another agent instead of from the low-level sensory environment. Hence, communication by an agent is inevitable for predicting its environmental state efficiently.

Learning with whom to communicate is crucial. Full communication does not scale well with the number of agents (Hoshen 2017). Predefined protocols cannot adapt to environmental changes or capture dynamic changes in agents' interactions (Han et al. 2018). Not all agents are equally informative in a situation (Kapourchali and Banerjee 2018a). Communication with a less-informative agent increases cost and might reduce the agent's confidence and accuracy.

Partially-observable Markov decision processes (POMDPs) have been widely used to learn a state-to-action mapping, referred to as *policy*, which requires a reward function dependent on the agent's goal. Predictive coding is a more general framework for modeling an agent, with no explicit reward function (Friston, Daunizeau, and Kiebel 2009; Banerjee and Dutta 2014). We propose an agent model in the predictive coding framework with a unified objective – minimization of variational free energy (VFE) – for inference, learning, and action. Using the same objective, our agent learns a communication policy as a mapping from its belief state to *with whom to communicate*.

Human activity recognition from multimodal, multisource and heterogeneous sensor data is used as a testbed to evaluate the proposed model. To test the model for larger number of agents, we use Kinect skeleton data from UTD-MHAD (Chen, Jafari, and Kehtarnavaz 2015) and KARD (Gaglio, Re, and Morana 2015) datasets where each joint in the skeleton is monitored by an agent. The learned policy is compared to a myopic policy as well as a decision-level fusion method where all agents send their messages to a central node. When all agents send reliable messages, an offline and myopic approach performs as good as the learned policy. However, when the probability of failure of each agent increases, online decision-making using the learned policy maintains the same accuracy by increasing the number of communications. If agents' behaviors change over time, the policy adapts to select other agents for communication.

The model is also applied to activity recognition from mutimodal UTD-MHAD dataset (Kinect skeleton, inertial and depth video). Each sensor is assumed to be monitored by an agent. A policy is learned for each activity class. Communication enhanced efficiency by using a subset of observations. The estimation accuracy is comparable to the state-of-the-art even though our model uses significantly fewer parameters and infers the state in a *localized* manner (i.e. it communicates neither with a central/global controller nor with all the agents all the time).

The rest of the paper is organized as follows. Sec. II introduces the necessary concepts. The problem statement and proposed model are described in Sec. III and IV, respectively. The experimental results are discussed in Sec. V. A brief literature review is provided in Sec. VI.

[1] Analysis of the properties of multiagent interaction to achieve a common goal using our proposed model, albeit interesting, is beyond the scope of this paper.

II. Background and Notations

This section introduces the relevant terms and concepts.

Table 1: Symbols and notations.

Variable	Description
I	Number of states.
J	Number of agents.
$\vec{\varphi}^{(e)} \in \Re^M$	Feature vector.
$\vec{\varphi}^{(msg)} \in \Re^I$	Communication message.
$\vec{\mu}^{(v)} \in \Re^I$	Belief vector about environmental states.
$\vec{\mu}^{(u)} \in \Re^J$	Belief vector about control states.
$\vec{\epsilon}_{\varphi^{(e)}} \in \Re^M$	Sensory prediction error.
$\vec{\epsilon}_{\varphi^{(msg)}} \in \Re^I$	Communication message prediction error.
$\vec{\epsilon}_{p^{(e)}} \in \Re^I$	Prior prediction error.
$\vec{v}_p \in \Re^I$	Mean of prior density.
$\mathbf{\Theta}_{g_e} \in \Re^{M\times I}$	Parameters for agent's model of environment and other agent $A_{j'}$ respectively.
$\mathbf{\Theta}_{g_{A_{j'}}} \in \Re^{I\times I}$	
$\mathbf{\Theta}_{g_\pi} \in \Re^{J\times I}$	Parameters for encoding optimal policy.
$\mathbf{\Sigma}_\chi$	Covariances of random fluctuations where $\chi = \{\vec{\varphi}^{(e)}, \vec{\varphi}^{(msg_{j'})}, \pi, p^{(e)}\}$.

Definition 1. (Agent) An agent is anything that can perceive its environment through sensors and act upon that environment through actuators (Russell and Norvig 2016). The agent estimating its environmental state will be referred to as the *primary agent*.

Definition 2. (Markov decision processes) Sequential decision problems in uncertain environments, also called Markov decision processes (MDPs) are defined as tuple (Russell and Norvig 2016): $\langle \Psi, A, T_a, r_a \rangle$ where Ψ is a finite set of states, A is a finite set of actions. $T_a(\psi'|\psi, a) = P(\{\Psi_{t+1} = \psi'|\Psi_t = \psi, A_t = a\})$ is the transition probability. r_a is the reward received at state ψ'. The goal is to find a policy $\pi : \Psi \rightarrow A$ that maximizes the cumulative rewards. The objective of MDP can be expressed as the Bellman optimality equation (Bellman 1952): $Value(\psi) = r_a + \max_{a\in A} \sum_{\psi'} T_a(\psi'|\psi, a) Value(\psi')$ where $Value(\psi)$ is the utility or value of state ψ.

Definition 3. (Partially observable MDPs) Partially observable MDPs (POMDPs) is an extension of MDP when the states are partially observable. A POMDP can be converted to a MDP using beliefs about the current state. The belief can be recursively computed from the observations and actions using Bayes rule. POMDP based approaches can provide a closed-loop non-myopic solution for agents' optimal decision-making problem (Russell and Norvig 2016).

Most of existing POMDP solvers are designed for purposes when reducing uncertainty is a subtask and not a goal. They fail for active perception due to requiring a long time for computing policy or underlying assumptions (e.g. piecewise linearity) that do not hold for a belief based reward function required for active perception (Satsangi et al. 2018).

Definition 4. (Predictive coding) Predictive coding (PC) is a brain-inspired framework for solving the problem of inferring the causes from sensations (Rao and Ballard 1999). Inspired by linearly solvable MDPs (Todorov 2007) and path integral control frameworks (Kappen, Gómez, and Opper 2012), a version of PC proposes an alternative approach for modeling an agent which is efficient and does not require a reward function to compute optimal policy (Friston, Daunizeau, and Kiebel 2009). By modeling action as inference and maximizing marginal likelihood of observations under a generative model, the optimal policy can be computed as a Kullback-Leibler (KL)-divergence minimization problem. A formal proof is provided in Friston, Daunizeau, and Kiebel (2009) to show that these *policies are equivalent to the ones computed using Bellman optimality equation (Def. 2). Hence PC is a generalization of optimal control or POMDPs.*

An agent in PC framework is defined as the tuple $\langle \Psi, A, \vartheta, G, Q, R, \Phi \rangle$ where Ψ is a set of states, A is a set of actions. ϑ is a set of real valued parameters. G and Q are generative and recognition densities. R is sampling probability and Φ is a set of sensory states (Friston, Samothrakis, and Montague 2012). The agent's objective is to minimize VFE which is a measure of salience based on the divergence between the recognition density $Q(\psi)$ and generative density $p(\varphi, \psi)$ (Friston, Daunizeau, and Kiebel 2009): $F = -\langle \ln p(\varphi, \psi) \rangle_Q + \langle \ln Q(\psi) \rangle_Q$ where $\langle . \rangle_Q$ denotes the expectation under density Q.

Definition 6. (Recognition density) Recognition density is a probabilistic representation of environmental states which is encoded by internal states μ. Probabilistic representation of environmental states is the agent's belief vector. Assuming a Gaussian density allows Laplace approximation (Friston, Daunizeau, and Kiebel 2009): $Q(\psi) = \mathcal{N}(\psi; \mu, \zeta) = \frac{1}{\sqrt{2\pi\zeta}} exp(-(\psi - \mu)^2/2\zeta)$. Sufficient statistics of a Gaussian density are mean and variance.

Definition 7. (Generative density) Generative density $p(\varphi, \psi)$ is a joint probability density relating environmental states and observations. It includes a sensory mapping $\varphi = g(\tilde{v}, \tilde{u}, \theta_g) + \tilde{\omega}_1$ and equation of motion $\dot{\tilde{v}} = f(\tilde{v}, \tilde{u}, \theta_f) + \tilde{\omega}_2$ (Friston, Daunizeau, and Kiebel 2009), where $\tilde{\omega}_i (i = 1, 2)$ are Gaussian noise. The latter contains the policies encoded in the parameters θ_f. It is a joint probability distribution over states, control states and the learned parameters. v and u are environmental hidden states and control states, respectively. $\tilde{X}$ shows the generalized coordinates of the variables. We use second order generalized coordinates consisting of state and change of state.

Definition 8. (Sampling probability) Sampling probability $R(\varphi'|\varphi, a) = p(\{\varphi_{t+1} = \varphi'|\varphi_t = \varphi, a_t = a\})$ is the probability that the observation $\varphi' \in \Phi$ follows action $a \in A$ given φ (Friston, Samothrakis, and Montague 2012).

III. Problem Statement

State estimation can be formulated as Bayesian inference (Knill and Richards 1996): $p(\Psi_t|\Phi_{1:t}) \propto p(\Phi_{1:t}|\Psi_t)p(\Psi_t)$. Active perception is defined as (Denzler and Brown 2002) $p(\Psi_t|A_{1:t}, \Psi_{1:t})$, in which the previous actions are causes for the current observation. Such problems are traditionally solved by POMDPs for non-myopic decision-making. We consider other agents as active parts of an agent's environment so that it can change its control states via communication which is an action. The problem is formulated as:

$$p(\Psi_t|A_{1:t},\Phi_{1:t}) = \frac{p(\Phi_{1:t}|\Psi_t,A_{1:t})p(\Psi_t,A_{1:t})}{p(\Phi_{1:t},A_{1:t})} \quad (1)$$

A number of challenges need to be addressed: (1) the size of action space grows exponentially with the number of agents, rendering standard POMDP solvers infeasible (Satsangi et al. 2018); (2) since all agents are not equally informative and their internal models are unobservable and time-varying, the problem needs to be solved online, without supervision or reinforcement; (3) an agent has to assign a degree of trust to each message received and update its belief accordingly.

IV. Models and Methods

We consider Ψ as a collection of causal environmental states that influences observations. It includes V as the uncontrollable aspects of environment and U which can be controlled by an agent. We model communication as an action using which an agent changes other agents' control states. We distinguish between A and U as an action may fail to control other agents. The action reveals a new observation, communication message $\Phi^{(msg)}$ that depends on U and V. Therefore, the random variable Φ collects two types of observations: $\Phi^{(e)}$ generated by the shared environment and $\Phi^{(msg)}$ generated by other agents as controllable parts of environment. The goal is to infer V at time t, efficiently, by activating the optimal sequences of $U_{1:t}$. Obviously, Φ_t is conditionally independent of action A, given Ψ which consists both U and V. Accordingly, the problem of *with whom to communicate* is converted to inferring the optimal sequence of control states $U_{1:t}$. Rewriting the above discussion as $p(\Psi_{1:t}|\Phi_{1:t})$, the problem is a Bayesian inference where exact computation is intractable for large distributions.

We approximate the posterior belief using variational inference (Fox and Roberts 2012), by minimizing divergence between a recognition density and the posterior density to reach $D_{KL}(Q(\Psi_{1:t})||p(\Psi_{1:t}|\Phi_{1:t})) = F + \ln p(\Phi_{1:t})$ where F is the VFE in Def. 4. Hence we can formulate our agent's model in the PC framework (Def. 4). We then provide an algorithm for sequentially optimizing perception and action, and updating agents' model as well as optimal policy. Ψ, Φ and A are defined above so rest of the elements are defined:

- ϑ represents real valued internal states of the agent which parameterize a conditional density.
- Generative density $G = p(\Phi_{1:t},\Psi_{1:t})$ relates environmental states and sensory data. It can be specified in the form of a likelihood and a prior. In our model, it is defined as: $p(\Phi_{1:t},\Psi_{1:t}) = p(\Phi_{1:t}|\Psi_{1:t})p(\Psi_{1:t})$ As in POMDPs, the Markovian assumption implies that Φ_t depends only on Ψ_t, so the likelihood term can be written as $p(\Phi_{1:t}|\Psi_{1:t}) = \prod_t p(\Phi_t|\Psi_t)$. The transition probabilities depend on the parameters ϑ. They are defined as: $p(\Psi_{1:t}) = p(\Psi_0)\prod_t p(\Psi_t|\Psi_{t-1},\vartheta)$. The prior expectations over trajectory of controlled states include policy (see Def. 7).
- Sampling probability $R = p(\Phi_{t+1}|\Phi_t,a_t)$ is agent's prediction of its action's consequences. The agent needs to learn an internal model of other agents to predict their responses to communication. The received message can be different from agent's prediction so the model is updated using prediction error.
- Recognition density $Q(\Psi_{1:t},\vartheta|\mu_{1:t})$, is an approximate posterior over states and parameters encoded with its sufficient statistic $\mu_{1:t}$, in the agent's internal model. The density is assumed to be Gaussian for Laplace approximation.

The unified objective of each agent for inference (perception), learning and communication (action selection in general) is to minimize the VFE (Def. 5).

Since $Q(\Psi_{1:t})$ is a Gaussian, with Laplace approximation, F converts to:

$$F = -\ln p(\mu_{1:t},\Phi_{1:t}) + C \quad (2)$$

where $\ln p(\mu_{1:t},\Phi_{1:t})$ is the generative density in which the environmental states are approximated by sufficient statistics of recognition density (agent's belief) and C is a constant which will be eliminated from rest of the equations for brevity. An intuitive interpretation of the above equation is that the agent interprets the external states of the environment (including both sensory states and hidden environmental states), in terms of its hidden internal states $\mu_{1:t}$. See (Buckley et al. 2017) for a formal proof.

A block diagram of our model in Fig. 1 provides an overview. Details of the blocks are as follows.

IV-A. Independent inference by an agent. In our model, an agent starts with an independent estimation based on its private sensory signals $\vec{\varphi}^{(e)}$. Vector sign indicates that the observation is multivariate. Since at this time only $\vec{\varphi}^{(e)}$ is available, the objective function is simplified to:

$$F^{(e)} = -\ln[p(\vec{\varphi}^{(e)}|\vec{\mu}^{(v)})p(\vec{\mu}^{(v)})] \quad (3)$$

where $p(\vec{\varphi}^{(e)}|\vec{\mu}^{(v)}) = p(\vec{\varphi}^{(e)}|\vec{v}) + \omega_1$ and $p(\vec{\mu}^{(v)}) = \vec{v}_p + \omega_2$. $\vec{\mu}^{(v)}$ denotes the belief vector regarding the aspect of environmental states $\vec{v}$, which should be estimated. Gaussian assumptions about error terms $w_i (i = 1, 2)$, specify likelihood and priors as $\mathcal{N}(\vec{\varphi}^{(e)}; g_e(\vec{\mu}^{(v)},\mathbf{\Theta}_{g_e}),\mathbf{\Sigma}_{\varphi^{(e)}})$ and $\mathcal{N}(\vec{\mu}^{(v)};\vec{v}_p,\mathbf{\Sigma}_{p^{(e)}})$, respectively. Mean of likelihood density, $g_e(\vec{\mu}^{(v)},\mathbf{\Theta}_{g_e}) = \mathbf{\Theta}_{g_e}\vec{\mu}^{(v)}$, is the generative function which maps agent's belief to the environmental observations $\vec{\varphi}^{(e)}$. In this paper, it is assumed to be a linear function, however, there is no limitation for using non-linear functions as long as they are differentiable. In our model, g_e is initialized using a limited number of samples and updated by observing each new sample in an online manner (details in Sec. VI). Plugging the Gaussians in Eq. 3, the best guess can be found by stochastic gradient descent:

$$\dot{\vec{\mu}}^{(v)} = \frac{\partial F^{(e)}}{\partial \vec{\mu}^{(v)}} = -\vec{\epsilon}_{p^{(v)}} + \frac{\partial g_e(\vec{\mu}^{(v)},\mathbf{\Theta}_{g_e})^T}{\partial \vec{\mu}^{(v)}}\vec{\epsilon}_{\varphi^{(e)}} \quad (4)$$

where $\vec{\epsilon}_{\varphi^{(e)}}$ and $\vec{\epsilon}_{p^{(v)}}$ are auxiliary variables representing $\mathbf{\Sigma}^{-1}_{\varphi^{(e)}}(\vec{\varphi}^{(e)} - g_e(\vec{\mu}^{(v)},\mathbf{\Theta}_{g_e}))$ and $\mathbf{\Sigma}^{-1}_{p^{(e)}}(\vec{\mu}^{(v)} - \vec{v}_p)$, respectively. These terms describe prediction errors weighted by

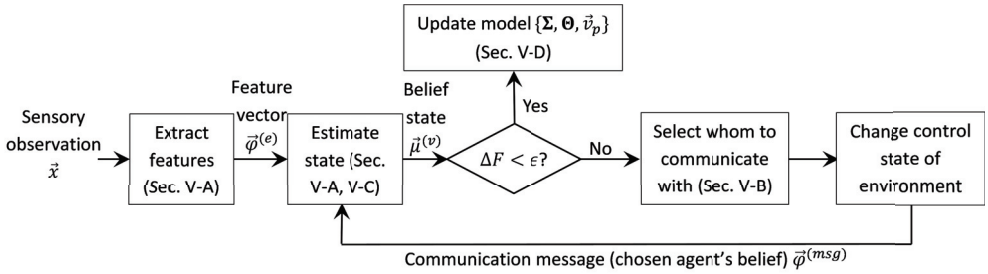


Figure 1: Block diagram of the proposed agent model for state estimation.

precision (inverse of variance). The former expresses deviation between agent's prediction $g_e(\vec{\mu}^{(v)}, \mathbf{\Theta}_{g_e})$ and actual observation $\vec{\varphi}^{(e)}$, while the latter denotes deviation of estimation $\vec{\mu}^{(v)}$ from prior expectation $\vec{v}_p$. Multiplying with precision terms weigh the influence of each error term in the inference. These weights define the relative degree of agent's attention to its prior knowledge and current sensory input.

IV-B. Selecting whom to communicate with. For each data sample, the agent ought to refine its initial and probably imprecise guess $\vec{\mu}^{(v)}$ through actions. Agents' actions change the control states of the environment, and hence the observations. Since communication is an action, the other agent's message will be an additional observation given that its control state is activated by the primary agent's action (request for communication). In this paper, we assume that the other agent sends its belief vector as the message. Taking into account the conditional independencies in our model, optimal action is selected as:

$$a_t = \underset{a}{\operatorname{argmin}} \sum_{\Phi} \underbrace{p(\vec{\varphi}^{(msg)}_{t+1}|\vec{\varphi}_t, a)}_{1} \Bigg[\ln \underbrace{p(\vec{\varphi}^{(e)}|\vec{\mu}^{(v)}_t)}_{2} + \sum_{\tau=1}^{t} \ln \underbrace{(\vec{\varphi}^{(msg)}_{\tau}|\vec{\mu}^{(u)}_{\tau}, \vec{\mu}^{(v)}_{\tau})}_{3} + \ln \underbrace{p(\vec{\mu}^{(v)}_t)}_{4} + \sum_{\tau=1}^{t} \ln \underbrace{p(\vec{\mu}^{(u)}_{\tau+1}|\vec{\mu}^{(u)}_{\tau}, \vec{\mu}^{(v)}_{\tau})}_{5} \Bigg] \quad (5)$$

where $\vec{\mu}^{(v)}_{t=1}$ is the agent's best guess calculated from Eq. 4. Eq. 5 implies agent A_j chooses to communicate with agent $A_{j'}$ ($a = j'$) whom A_j believes would maximally decrease the VFE. The second and fourth terms are defined in the last section, following Eq. 3. The third term contains model of another agent. An agent needs to learn a model of other agents from their messages, in order to interpret the observations generated by them. This model has the same form as the generative function of environment g_e but with different parameters: $\mathcal{N}(\vec{\varphi}^{(msg_{j'})}; g_{A_{j'}}(\vec{\mu}^{(v)}, \vec{\mu}^{(u)}, \mathbf{\Theta}_{g_{A_{j'}}}), \mathbf{\Sigma}_{\varphi^{(msg_{j'})}})$ where $g_{A_{j'}}(\vec{\mu}^{(v)}, \vec{\mu}^{(u)}, \mathbf{\Theta}_{g_{A_{j'}}}) = \mu^{(u_{j'})}\mathbf{\Theta}_{g_{A_{j'}}}\vec{\mu}^{(v)}$ where $\mu^{(u_{j'})} = 1$ means that control state of $A_{j'}$ is activated by action. The parameters $\mathbf{\Theta}_{g_{A_{j'}}}$, are learned over time by the samples of communication provided by $A_{j'}$ to A_j and are unique for each agent in the environment.

The fifth term represents agent's prior beliefs about transition among states. It depends on the parameters ϑ. Optimal priors over these parameters make this term equivalent to optimal policy (Friston, Samothrakis, and Montague 2012). In other words, $p(\vec{\mu}^{(u)}_{\tau+1}|\vec{\mu}^{(u)}_{\tau}, \vec{\mu}^{(v)}_{\tau}) = T(\Psi_{\tau+1}|\Psi_{\tau}, \pi(\Psi_{\tau})) + \omega_3 = T(U_{\tau+1}|U_{\tau}, V_{\tau}, \pi(\Psi_{\tau})) + \omega_3$, where V_{τ} does not change over $\Delta\tau \rightarrow 0$ so $V_{\tau+\Delta\tau} \approx V_{\tau}$. Therefore, the fifth term is a Gaussian $\mathcal{N}(\vec{\mu}^{(u)}_{\tau+1}; g_{\pi}(\vec{\mu}^{(u)}_{\tau}, \vec{\mu}^{(v)}_{\tau}, \mathbf{\Theta}_{\pi}), \mathbf{\Sigma}_{\pi})$. In this paper, the next control state $\vec{\mu}^{(u)}_{\tau+1}$ needs to be inferred since the agent should choose the communication target. The agent knows with whom it has already communicated so $\vec{\mu}^{(u)}_{\tau} = \vec{u}_{\tau}$. Thus it will communicate with $A_{j'}$ if $\mu^{(u_{j'})}_{\tau} = u_{j'} = 0$. The generative function for trajectory of control states (priors on the dynamics) is defined as: $g_{\pi}(\vec{\mu}^{(u)}, \vec{\mu}^{(v)}, \mathbf{\Theta}_{\pi}) = (\vec{\mathbf{1}} - \vec{\mu}^{(u)}) \odot (\mathbf{\Theta}_{\pi}\vec{\mu}^{(v)})$ where $\vec{\mathbf{1}} \in \Re^J$ and $\odot$ is element-wise product. Finally, the first term in Eq. 5 is the sampling probability. It allows the agent to predict other agents' behaviors given the current evidences. $\vec{\varphi}^{(msg)}_{t+1}$ is A_j's prediction about the next observation.

IV-C. Updating belief using communication message. The received communication message ($\vec{\varphi}^{(msg)}_{t+1}$) is a new observation. It is interpreted through agent's internal model in the same way $\vec{\varphi}^{(e)}$ is processed. This helps the agent to *reason whether it wants to update its belief or not* based on the reliability of the sender. Reliability of $A_{j'}$'s messages are measured by the precision term, $\mathbf{\Sigma}^{-1}_{\varphi^{(msg_{j'})}}$. The agent's belief is updated by minimizing $F(\{\vec{\varphi}^{(e)}, \vec{\varphi}^{(msg)}_1, ..., \vec{\varphi}^{(msg)}_{t+1}\})$:

$$\dot{\vec{\mu}}^{(v)}_{t+1} = \frac{\partial F}{\partial \vec{\mu}^{(v)}_{t+1}} = -\vec{\epsilon}_{p^{(v)}} + \frac{\partial g_e(\vec{\mu}^{(v)}_{t+1}, \mathbf{\Theta}_{g_e})^T}{\partial \vec{\mu}^{(v)}_{t+1}}\vec{\epsilon}_{\varphi^{(e)}} + \sum_{\tau=1}^{t+1} \frac{\partial g_{A_{j'}}(\vec{\mu}^{(v)}_{\tau}, \vec{\mu}^{(u)}_{\tau}, \mathbf{\Theta}_{g_{A_{j'}}})^T}{\partial \vec{\mu}^{(v)}_{\tau}}\vec{\epsilon}_{\varphi^{(msg)}_{\tau}} + \sum_{\tau=1}^{t+1} \frac{\partial g_{\pi}(\vec{\mu}^{(u)}_{\tau}, \vec{\mu}^{(v)}_{\tau}, \mathbf{\Theta}_{\pi})^T}{\partial \vec{\mu}^{(v)}_{\tau}}\vec{\epsilon}_{\pi} \quad (6)$$

where $\vec{\epsilon}_{\pi} = \mathbf{\Sigma}^{-1}_{\pi}(\vec{\mu}^{(u)} - g_{\pi}(\vec{\mu}^{(u)}, \vec{\mu}^{(v)}, \mathbf{\Theta}_{\pi}))$ and $\vec{\epsilon}_{\varphi^{(msg)}} = \mathbf{\Sigma}^{-1}_{\varphi^{(msg_{j'})}}(\vec{\varphi}^{(msg_{j'})} - g_{A_{j'}}(\vec{\mu}^{(v)}, \vec{\mu}^{(u)}, \mathbf{\Theta}_{g_{A_{j'}}}))$. Since now

$t+1$ is the current time, $\varphi_{t+1}^{(msg)}$ is the observation and not a prediction.

IV-D. Updating the agent's internal model. Updating the model, in an online and unsupervised manner, helps the agent to progressively adapt itself to minimize VFE on successive exposure to the same stimulus. In our model, after each communication sequence, if the VFE has converged, the agent updates its model. Here we provide the update rules for parameters and hyperparameters of the model. Parameters of g_e are updated as:

$$\frac{\partial F}{\partial \mathbf{\Theta}_{g_e}} = \mathbf{\Sigma}_{\varphi^{(e)}}^{-1}(\vec{\varphi}^{(e)} - g_e(\vec{\mu}^{(v)}, \mathbf{\Theta}_{g_e}))\vec{\mu}_T^{(v)^T} = \vec{\epsilon}_{\varphi^{(e)}}\vec{\mu}_T^{(v)^T}$$

where superscript T refers to the matrix transpose operation while subscript T stands for the total communication time (i.e. $T = J$ or total number of agents communicated with when $\Delta F < \epsilon$). Model of agent A_j from each agent $A_{j'}$ where $j' \in \{1, ...J\}$ and $j' \neq j$ is updated as:

$$\frac{\partial F}{\partial \mathbf{\Theta}_{g_{A_{j'}}}} = \vec{\epsilon}_{\varphi^{(msg_{j'})}}\vec{\mu}_T^{(v)^T} \tag{7}$$

Parameters of optimal policy after taking each action at time t, where $\mu_t^{(u_{a_{t-1}})} = 1$, is updated as:

$$\frac{\partial F}{\partial \mathbf{\Theta}_\pi} = (1 - \vec{\mu}_{t-1}^{(u)}) \odot \vec{\epsilon}_\pi \vec{\mu}_{t-1}^{(v)^T} \tag{8}$$

The update rules for covariance matrices are:

$$\frac{\partial F}{\partial \mathbf{\Sigma}_\chi} = \frac{1}{2}(\vec{\epsilon}_\chi \vec{\epsilon}_\chi^T - \mathbf{\Sigma}_\chi^{(-1)}) \tag{9}$$

where χ is replaced with the $\vec{\varphi}^{(e)}$, $\vec{\varphi}^{(msg_{j'})}$, π and $p^{(e)}$.

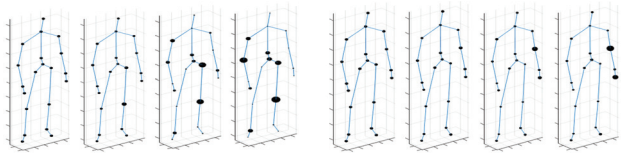
V. Experimental Results

The model is evaluated for human activity recognition. Sensors generate high-dimensional multivariate time-series. We use a convolutional sparse coding model (Kapourchali and Banerjee 2018b) to learn a dictionary of features from data. The sequence of indices of the detected feature ($o_{\vec{d}}^m$) and corresponding shift (o_τ^m) for each variable constitutes the sensory feature vector (the agent's sensory observation): $\vec{\varphi}^{(e)} = [o_{\vec{d}}^1, o_\tau^1, o_{\vec{d}}^2, o_\tau^2, ..., o_{\vec{d}}^{M'}, o_\tau^{M'}]^T$, where M' is the number of variables. Our model is applied to two experiments: (1) skeleton-based activity recognition to evaluate the model for larger number of agents, and (2) multimodal activity recognition to evaluate the model on heterogenous data.

Skeleton-based human activity recognition. Benchmark datasets for activity recognition rarely exceed a few sensors. So the model is evaluated on two benchmark datasets for activity recognition using Kinect skeleton data where each joint is assumed to be monitored by an agent. KARD dataset (Gaglio, Re, and Morana 2015) comprises of 18 activities performed by 10 individuals. Each person repeated each activity three times. The dataset includes 540 sequences. The skeleton has 15 3-D joints. UTD-MHAD (Chen, Jafari, and Kehtarnavaz 2015) is a multimodal dataset which also includes skeleton data. It has 27 activities performed by eight subjects. Each subject performed each activity four times. After removing three corrupted sequences, the dataset includes 861 sequences. The skeleton has 20 3-D joints. Each agent observes only its 3-D signals and the communication messages from other joints (agents) upon request. It does not have access to the observations and internal models of other agents. In order to compare with baselines, the "new person" setup, as in Gaglio, Re, and Morana (2015), is used where data of one subject is reserved for testing while the model is trained on data of other subjects.

First, a dictionary of 50 features is learned from the training set. Inference starts with the head (primary) agent (the joint representing the head of the person) though this does not have to be the case. From the index of the best matched feature for each of the three coordinates and their corresponding optimal shifts, the posterior probability distribution over all possible states (activity categories) is inferred by the primary agent, independently. The agent iteratively refines the belief using the steps shown in the Fig. 1. Communication stops if the change in VFE is less than ϵ ($= 10^{-3}$). The internal model of the primary agent is updated based on the final inference.

Fig. 2 shows the learned policies for a particular subject for two activity classes. The learned policy for different activities are different. The head agent relies on the agents located in parts of the body with more variations in the environmental signals during that activity. Fig. 3 shows the final learned policy for a situation where the head agent fails to distinguish between two activities: *Lunge* and *Bowling*. The head agent inferred *Lunge* as *Bowling* half of the times. A sample frame of a subject's posture for each of these activities are shown. The largest circle belongs to the wrist agent (hand agent is not visible in the figure due to its small size). Based on information theory, it is expected that the head agent chooses the agents in the most salient parts of the body during a particular activity (i.e. the signals with less mutual information) (Russell and Norvig 2016). Saliency of an agent is measured by the KL-divergence between its belief distribution and that of the head (primary) agent's.



(a) Walking (b) Waving

Figure 2: The learned policies for two activity classes. Number of training iterations (from left to right): 1, 100, 500, 1000. Length of a circle's radius is proportional to the probability of communicating with the corresponding joint-agent.

Fig. 3(b) compares the saliency of different agents' beliefs. A circle's radius is proportional to KL-divergence between distributions. However, this saliency is with respect to the head (primary) agent at the initial step without considering the pairwise similarity between the beliefs of other agents. Two agents might convey the same information so that once the head agent communicates with one of them, the

other one is no longer salient. A non-myopic approach takes the conditional saliency into account. It can be seen that only a subset of the most salient joints are in the learned policy. To visualize this, we grouped the agents' beliefs using k-means clustering and plotted the joints in the same cluster with the same color. The number of clusters is decided based on average number of times the agents communicated for this activity class. The silhouette coefficients indicate the clusters are reasonably compact and homogeneous (ref. Fig. 3(c)). Even though the saliency of the hip-center agent is less than some of the others, in the policy distribution it has a higher weight because it is alone in its cluster and no other agent's belief is similar to its. Among the more salient joints, at least one from each cluster is present in the learned policy.

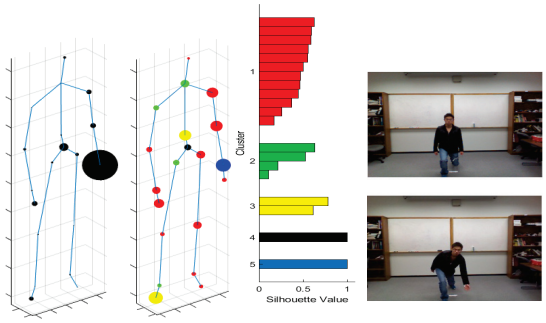


Figure 3: (a) Policy when desired state is *Lunge* but the head agent infers *Bowling* from its environmental observations. (b) Saliency of each joint (colors show clusters). (c) Silhouette coefficient. (d) A sample frame from each activity.

Fig. 4 shows an example of sequential decision-making by an agent for whom to communicate with. It shows how the head agent decides on a sequence of actions to decrease the uncertainty. We have intentionally chosen an activity regarding which the head agent is highly uncertain and ends up communicating with six other agents before reaching the final decision. The activity is *Knocking*. First, the head agent infers it as *Jogging*. Refer to the first top left subfigure in Fig. 4 and the corresponding belief. This belief has high entropy, so the agent communicates with the wrist agent to reduce uncertainty. It can be seen that the maximum belief is changed to the 21st activity which is *Pick up and Throw* (note that throwing involves wrist movement similar to knocking). The communication continues by requesting belief from the hip agent. It reduces the uncertainty in belief by decreasing the second maximum probability. That is, by asking the hip agent, the agent recognizes the activity is not *Lunging*. Finally, the agent reaches the correct state by communicating with shoulder agent and becomes more certain by communicating with elbow and shoulder center agents.

For quantitative evaluation, two cases are considered: (1) the probability of each agent sending random responses is non-zero, and (2) a fixed set of agents, drawn from a uniform distribution, generate random beliefs for a number of trials. We compare our model with two widely-used decision-making methods: (1) an information theoretic technique, Value of Information (ref. Chapter 16 of (Russell and Norvig 2016)), as a myopic and offline decision-making, and (2) fusion where the posterior probability is computed at a central node as weighted mean of all agents' beliefs. Results are shown in Fig. 5. When agents randomly fail to provide informative messages, online non-myopic decision-making helps to maintain accuracy by increasing the number of communications. However, when the same agents fail to send informative messages for a long time, updating the agents' models helps the primary agent to adapt its policy; the increase in number of communications is less compared to a non-adaptive approach.

Table 2 shows the head agent's inference accuracy using different communication protocols. Using learned policy, the accuracy of recognition is increased. The head agent communicated 63.54% of the time for KARD and 61.32% of the time for UTD-MHAD dataset which is a significant saving in time and resources. Accuracies from references are provided as a baseline. Note that the accuracy of our model also depends on the nature of the chosen generative function, number of parameters, and dimension of hidden state vector. Accuracy can be improved by replacing our linear generative function with a more sophisticated one.

Table 2: Recognition accuracy(%) for the two datasets. "No Comm" and "Full Comm" refer to accuracy of the agent alone and when the agent communicates with *all* other agents. "Policy" refers to our model. "Ref." provides baseline accuracy for the new person setup in (Gaglio, Re, and Morana 2015) and (Chen, Jafari, and Kehtarnavaz 2016) for KARD and UTD (Kinect alone).

	No Comm	Full Comm	Policy	Ref.
KARD	24.2±1	88.1±1	90.2±3	84.6
UTD-MHAD	18.6±2	73.1±4	80.1±4	74.7

Multimodal human activity recognition. The proposed model is evaluated for multimodal activity recognition on UTD-MHAD dataset which is introduced in the last section where only Kinect skeleton was used. In this section, data from different modalities, namely, depth, skeleton and inertia are used where each sensor modality is assumed to be monitored by an agent. The frame size in depth data is reduced by a factor of 10 to enhance depth agent's efficiency. The agents' generative functions are learned using data from four subjects (subjects 1 through 4) which are excluded from rest of the experiments. These subjects are considered in (Chen, Jafari, and Kehtarnavaz 2015) as training set, so using them for training allows appropriate comparison.

The inertia (primary) agent starts the communication process since it has the least number of variables (three variables leading to a 6-D feature vector) which incurs lower computational cost. After an independent inference, it communicates with an agent based on the optimal policy and decides to further communicate until the convergence criterion is satisfied. Recognition accuracy for different kinds of communication are shown in the bottom three rows of Table

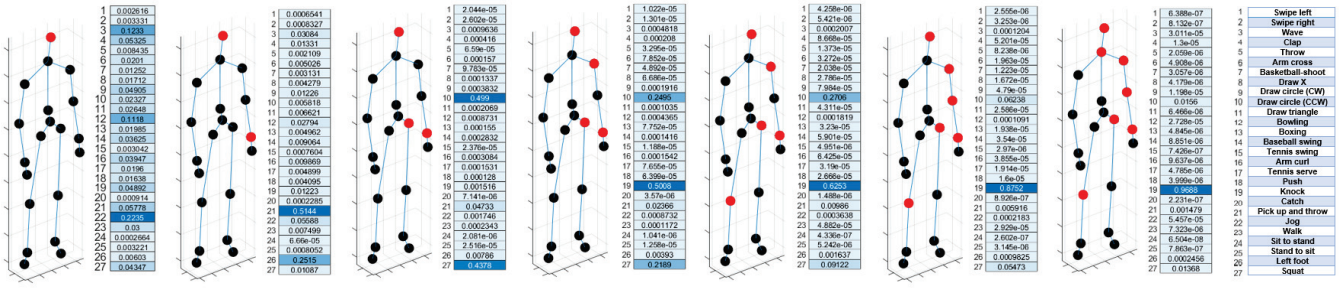


Figure 4: Sequential decision-making for *with whom to communicate*. Red circle denotes the agent $A_{j'}$ selected for communication. Primary agent A_j's belief vector (probability of each environmental state or activity) after communication is shown.

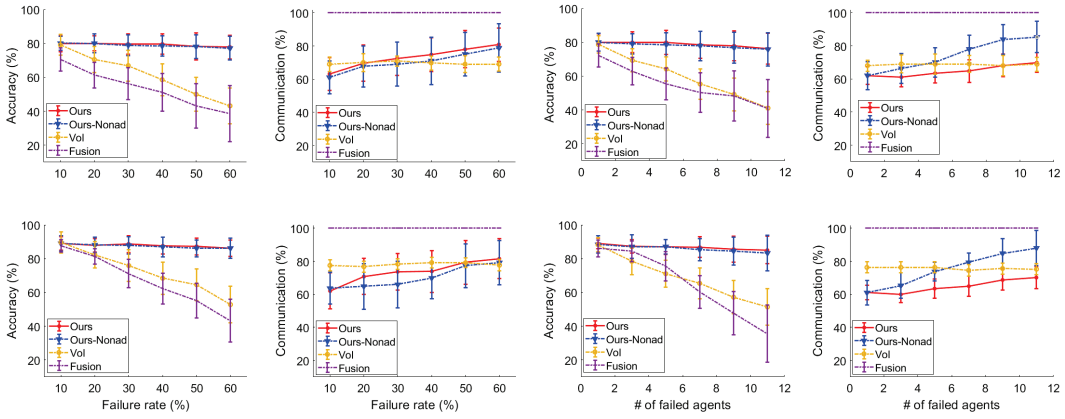


Figure 5: Advantages of online, non-myopic decision-making, as well as online updating of agents' model are shown in these figures. Top and bottom rows are the results from UTD-MHAD and KARD datasets, respectively. The two left plots from each row shows the accuracy and number of communications when each agent has a probability of failure at each point of time. The two right plots from each row show the same metrics but a fixed number of agents, sampled from a uniform distribution, change their behavior and send random messages for a long time. Nonad and VoI stand for Non-adaptive and Value of Information (a myopic and offline planning method) methods, respectively.

3. Results show the benefit of communication. However, full communication does not guarantee highest accuracy.

Our model is compared with existing methods that have used the same cross-subjects setup for training. The results show that even though our model has significantly fewer parameters, communicating using a learned policy yields higher accuracy than most of these methods (see Table 3). ConvNets (Hou et al. 2016) is slightly (1.86%) more accurate than our model; it has 60×10^6 parameters as compared to 67×10^3 in our model. The inertia agent communicated for 301 and 129 of the test samples with skeleton and depth respectively, but only three times with both.

VI. Related Work

Prior work on active perception has primarily focused on one agent controlling its sensors (Butko and Movellan 2009) or selecting a subset of sensors (Li et al. 2016; Satsangi et al. 2018). Research has been reported on controlling multiple sensors in which, whom to communicate with is either predefined (Zivan et al. 2015; Kapourchali and Banerjee 2019) or decided by a fusion center (Stachura and Frew 2017). In other areas, such as distributed AI and multiagent systems, some recent works (Hoshen 2017) have investigated the importance of learning with whom to communicate where the goal is coordination between agents. They use a single network for controlling a multiagent system (i.e. communication policies are globally learned) and lack the ability to handle heterogeneous agent types (Peng et al. 2017). In our model, policy is learned and executed locally; the task is active perception. Challenges of policy learning for such a task are discussed in (Satsangi et al. 2018).

VII. Conclusions

We propose an agent model for efficiently predicting its environmental state via selective communication with other agents. The agent is modeled in the predictive coding framework. It learns a communication policy as a mapping from its belief state to *with whom to communicate* in an online and unsupervised manner, without any reinforcement. The proposed model is evaluated for activity recognition from mul-

Table 3: Comparison of proposed and existing methods for recognizing 27 actions in the UTD-MHAD dataset.

Method	Accuracy %
ELC-KSVD (Zhou et al. 2014)	76.19
Chen, Jafari, and Kehtarnavaz (2015)	79.10
Cov3DJ (Hussein et al. 2013)	85.58
ConvNets (Hou et al. 2016)	86.97
Dawar and Kehtarnavaz (2018)	86.3
Our model	**85.11**
No Comm	29.2
Full Comm	84.6

timodal, multisource and heterogeneous sensor data. The accuracy is comparable to the state-of-the-art even though our model uses significantly fewer parameters and infers the state in a localized manner. The learned policy reduces number of communications and enhances tolerance to communication failures. To the best of our knowledge, this is the first work on learning communication policies by an agent for predicting the state of its environment.

References

Bajcsy, R.; Aloimonos, Y.; and Tsotsos, J. K. 2018. Revisiting active perception. *Autonomous Robots* 42(2):177–196.

Banerjee, B., and Dutta, J. K. 2014. SELP: A general-purpose framework for learning the norms from saliencies in spatiotemporal data. *Neurocomputing* 138:41–60.

Bellman, R. 1952. On the theory of dynamic programming. *PNAS* 38(8):716–719.

Buckley, C. L.; Kim, C. S.; McGregor, S.; and Seth, A. K. 2017. The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*.

Butko, N. J., and Movellan, J. R. 2009. Optimal scanning for faster object detection. In *CVPR*, 2751–2758. IEEE.

Chen, C.; Jafari, R.; and Kehtarnavaz, N. 2015. Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In *ICIP*, 168–172. IEEE.

Chen, C.; Jafari, R.; and Kehtarnavaz, N. 2016. A real-time human action recognition system using depth and inertial sensor fusion. *IEEE Sensors Journal* 16(3):773–781.

Dawar, N., and Kehtarnavaz, N. 2018. Real-time continuous detection and recognition of subject-specific smart tv gestures via fusion of depth and inertial sensing. *IEEE Access* 6:7019–7028.

Denzler, J., and Brown, C. M. 2002. Information theoretic sensor data selection for active object recognition and state estimation. *IEEE Trans. PAMI* 24(2):145–157.

Fox, C., and Roberts, S. 2012. A tutorial on variational bayesian inference. *Artif. Intell. Rev.* 38(2):85–95.

Friston, K.; Daunizeau, J.; and Kiebel, S. 2009. Reinforcement learning or active inference? *PloS one* 4(7):e6421.

Friston, K.; Samothrakis, S.; and Montague, R. 2012. Active inference and agency: optimal control without cost functions. *Biological cybernetics* 106(8-9):523–541.

Gaglio, S.; Re, G. L.; and Morana, M. 2015. Human activity recognition process using 3-d posture data. *IEEE Trans. Human-Mach. Syst.* 45(5):586–597.

Han, X.; Yan, H.; Zhang, J.; and Wang, L. 2018. Acm: Learning dynamic multi-agent cooperation via attentional communication model. In *ICANN*, 219–229. Springer.

Hoshen, Y. 2017. Vain: Attentional multi-agent predictive modeling. In *NIPS*, 2701–2711.

Hou, Y.; Li, Z.; Wang, P.; and Li, W. 2016. Skeleton optical spectra based action recognition using convolutional neural networks. *IEEE Trans. Circuits Syst. Video Technol.*

Hussein, M. E.; Torki, M.; Gowayyed, M. A.; and El-Saban, M. 2013. Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations. In *IJCAI*, volume 13, 2466–2472.

Kapourchali, M. H., and Banerjee, B. 2018a. Multiple heads outsmart one: A computational model for distributed decision making. In *CogSci*, 1779–1784.

Kapourchali, M. H., and Banerjee, B. 2018b. Unsupervised feature learning from time-series data using linear models. *IEEE Internet Things J.* 5(5):3918–3926.

Kapourchali, M. H., and Banerjee, B. 2019. State estimation via communication for monitoring. *IEEE Trans. Emerg. Topics Comput. Intell.*

Kappen, H. J.; Gómez, V.; and Opper, M. 2012. Optimal control as a graphical model inference problem. *Machine learning* 87(2):159–182.

Knill, D. C., and Richards, W. 1996. *Perception as Bayesian inference*. Cambridge University Press.

Li, Y.; Jha, D. K.; Ray, A.; and Wettergren, T. A. 2016. Sensor selection for passive sensor networks in dynamic environment: A dynamic data-driven approach. In *Am. Control Conf.*, 4924–4929. IEEE.

Peng, P.; Yuan, Q.; Wen, Y.; Yang, Y.; Tang, Z.; Long, H.; and Wang, J. 2017. Multiagent bidirectionally-coordinated nets for learning to play starcraft combat games. *arXiv preprint arXiv:1703.10069*.

Rao, R. P., and Ballard, D. H. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2(1):79.

Russell, S. J., and Norvig, P. 2016. *Artificial intelligence: A modern approach*. Malaysia; Pearson Education Limited.

Satsangi, Y.; Whiteson, S.; Oliehoek, F. A.; and Spaan, M. T. 2018. Exploiting submodular value functions for scaling up active perception. *Autonomous Robots* 42(2):209–233.

Stachura, M., and Frew, E. 2017. Communication-aware information-gathering experiments with an unmanned aircraft system. *Journal of Field Robotics* 34(4):736–756.

Todorov, E. 2007. Linearly-solvable markov decision problems. In *NIPS*, 1369–1376.

Zhou, L.; Li, W.; Zhang, Y.; Ogunbona, P.; Nguyen, T.; and Zhang, H. 2014. Discriminative key pose extraction using extended lc-ksvd for action recognition. In *IEEE DICTA*, 1–8.

Zivan, R.; Yedidsion, H.; Okamoto, S.; Glinton, R.; and Sycara, K. 2015. Distributed constraint optimization for teams of mobile sensing agents. *AAMAS* 29(3):495–536.