Communication as moving target tracking

Dynamic Bayesian inference with an action-perception-learning cycle

Byoung-Tak Zhang

We view communication partners as moving targets. Achieving the goal of communication, thus, requires tracking the conversational trajectory of the partner in real time. We formalize this as dynamic inference with an action-perception-learning cycle and use sequential Bayesian estimation to do that. Our information-theoretic, dynamic Bayesian formulation suggests to understand communication as a Markov decision process where one participant tries to simultaneously improve predictions about its partner's future state, manipulate the partner into states that maximize predictive information and minimize decision costs. The dynamic inference cycle model offers an overarching framework in which the mathematical tools developed in different fields can be used for modelling communication. It also helps develop technologies for multimodal embodied interaction and human-like cognitive agents.

1. Computational modelling of human communication

How can human communication be so flexible, fast, and still robust? How can we build agents that interact with humans as in human-human communication? Despite the remaining challenges to cognitive science and artificial intelligence, the last two decades of research gives us partial answers to these questions. Psycholinguists have discovered how the interlocutors align their representations to adapt their communicative behaviour to that of their conversational partner (Brennan et al., 2010). Cognitive scientists have figured out how the mind and body work together for effective communication (Sebanz et al., 2006; Knoeferle & Crocker, 2007; Spivey, 2007). Neuroscientists have identified the predictive mechanisms in the brain to perform computations related with perception, action, and cognition (Gallese et al., 1996; Friston, 2009). Computer scientists can now simulate some essential aspects of human communication, such as gestures and facial expressions (Kopp & Wachsmuth, 2010).

One of the main challenges in computational modelling of human communication is that the communicative process is highly dynamic. This makes the traditional rule-based modelling approach ineffective (Langley et al., 2009). Design of rules for the dynamic process is practically impossible. Recently, a variety of corpora on real-life conversations has been collected and a new design methodology has emerged. These so-called data-driven methods rely on machine learning techniques. Supervised machine learning has been especially successful in many problem domains in which labeled data are available. However, natural human communication requires a long-term sequence of multidimensional temporal data for which detailed manual labeling is infeasible.

We aim to resolve this situation by presenting a predictive model of human communication that is flexible and learnable from unlabeled conversation data. We formulate the communication as a problem of moving target tracking, where the target is the communication partner. This allows us to apply existing mathematical tools such as control theory, information theory, statistical physics, and computational learning theory. We develop a dynamic Bayesian inference framework that learns the target trajectory sequentially and online as communication unfolds in real time. The key assumption is that the agent is equipped with sensors and motors to monitor the partner (environment) and itself to get learning feedback in the perception action cycle (Fuster, 2001). For humans this assumption is naturally met. For artificial agents communicating with a human with speech and gesture, the requirement can be met by being equipped with cameras, eye trackers, and sound sensors. Embodied with many sensors, the agent can have a significantly enhanced autonomy of communication. Therefore, we also discuss communication situations where the agent receives rewards explicitly (from the partner) and implicitly (by internally-generated feedback).

The chapter is organized as follows. In Section 2 we describe the moving target analogy of communication. Section 3 formulates the dynamic inference cycle model of human communication. Section 4 discusses how the model can be learned automatically based on explicit and implicit feedback from the communication partner. Section 5 discusses how such a model can be applied further to human-like conversational agents, multimodal embodied interaction of robots, and towards human-level AI.

2. Communication as tracking moving targets

2.1 Entrainment and alignment

Mounting evidence suggests that communication is an interactive adaptation process, such as entrainment and alignment. In conversation, people tend to adapt their communicative behaviour to that of their conversational partner. Entrainment is a tendency of interlocutors to become similar, synchronous, and convergent (Brennan et al., 2010). Entrainment occurs in gesture, posture, and facial expressions as well as lexical, phonological, and syntactic aspects (Inden et al., 2012). Entrained interactions are perceived more attractive and intimate. Entrainment also makes the interactions more successful. Communication also involves with alignment, i.e. a process of determining correspondences between concepts or linguistic units.

A fundamental issue in communication is to understand the intention of the conversation partner (De Ruiter & Cummins, in press). Intention recognition involves managing the common ground, which is computationally very complex. Priming has been suggested as a method humans use to reduce the computational overhead bottleneck (Pickering & Garrod, 2004). Interlocutors can prime and share their representations between them. This is a fast, automatic process requiring low cognitive overhead. People also use situational contexts and conventions to recognize the intention of the partner. In discourse, statements are typically followed by agreement or disagreement. Invitations are followed by acceptance or declining. Social conventions can speed up processing significantly.

De Ruiter & Cummins (in press) propose a Bayesian model of intention recognition. They use a probabilistic modular approach where inputs and outputs are discrete probability distributions that are suitable for processing by other modules. It uses context, conventions, and likelihood to enhance efficiency of communication. Intention recognition is a mapping problem, i.e. maps linguistic and other signals into intension at time *t*. Thus, its emphasis is not on language production.

Pickering and Garrod (in press) present an integrated model of language production and comprehension. They present an alignment mechanism between speakers and listeners which is based on interweaving between processes of language production and comprehension within each interlocutor. Specifically it is argued that actors construct forward models of their actions before they execute those actions, and that perceivers of others' actions covertly imitate those actions, then construct forward models of those actions. It remains how the presented model can be realized in computational systems.

Modelling the alignment and entrainment processes in embodied communication (Allwood, 2008) offers several challenges. It requires multi-disciplinary approaches from psychology, linguistics, phonetics, and computer science that deal with facial expressions, spoken dialogue, eye tracking, gesture analysis, and other sensorimotor signals. From computer science point of view, this requires real-time interactive modelling of multidimensional spatiotemporal data over a long period of time. The dynamics and structure of the data is a real challenge even to the modern machine learning techniques. During the last decades, a variety of corpora have been collected, but most of these are from controlled situations and thus have ecologically-limited reality. Also previous studies focused on supervised learning formulations, ignoring the temporal dynamics of the entrainment in communication.

2.2 Communication and target tracking

In this article we focus on the long-term sequential and dynamic nature of entrainment rather than the local short-term aspects of conversation. To take the global view of conversation, we consider communication as tracking a moving target. In this view, the communication partner is a moving object like a cruise missile or flying airplane. The agent is like radar that transmits (acts) radio signals to the unknown target and receives (perceives) the returning signals. The agent has to estimate the state of the unknown target located in the environment by observing a sequence of radio signals. Notice that in this analogy the agent is active and autonomously decides which radio signals to be sent to which positions. The transmitted radio signals are like the utterances the agent produces. The returning signals are the utterances of the communication partner. Notice also that the agent continually updates the state of the target, such as the position and velocity, in the airplane example. Likewise the agent should continually attempt to recognize the belief, desire, and intention (BDI) of the communication partner.

In a more general case, the moving target can be a dancing partner with which the agent has to collaborate according to some protocol or conventions. Achieving the goal of communication, thus, requires tracking the movement trajectory by a continuous update cycles of prediction and correction in real time. Of course, human communication is not exactly like a flying airplane or a dancing partner. However, the analogy is very useful since it allows us to start the challenging computational modelling with many mathematical techniques developed in control theory and dynamical systems theory. As we shall see in Section 4 below, the analogy facilitates the use of other mathematical tools in information theory, statistical mechanics, and machine learning.

In the next section we formalize the target tracking as dynamic inference in which the agent sequentially produces utterances (actions) to test the "evolving" communicative goals of the partner in situ and in context. The utterances are chosen from the prior distribution of the hypotheses (past) and acted on the partner. The partner's responses are perceived to estimate the likelihood of the hypotheses (present). Then, the hypotheses are revised and learned to produce the posterior distribution (future). The updated hypotheses then form the prior distribution for the next cycle of Bayesian predictive inference. Here we develop a general framework for communication that constantly learns over an extended period of time.

We start by considering the information flow in the perception-action cycle of an agent interacting with the partner (environment).

3. The dynamic inference cycle in human communication

3.1 The Action-perception-learning cycle

Consider an agent in interaction with an environment (conversation partner) (Figure 1). The agent has a memory to model the conversation history. We denote the memory state at time t by m_t. The agent observes the environment and measures the sensory state s_t of the environment (perceives the partner's utterance) and chooses an action a_t. Note that the memory states are not observable while the world states are observable. The goal of the agent is to learn about the environment and predict the next world states s_{t+1} as accurately as possible. Note that the perception-action cycle of the agent models the continuous interaction with the partner in sequence.

In other problem settings, the agent can have a specific goal, such as completing a mission as in making reservations for a flight ticket through a phone call. In this case, the agent can receive rewards depending on its performance, e.g. success or failure of the mission and the length of conversation to complete the mission. From the actions a_t taken at state s_t and the rewards r_t received from the environment, the objective of the agent can be formulated to choose the actions that maximize the expected reward or value $V(s_t)$ in the future. Markov decision problems (Sutton & Barto, 1998) are a representative example of this class of tasks. We shall discuss these variants of objective functions in Section 4.

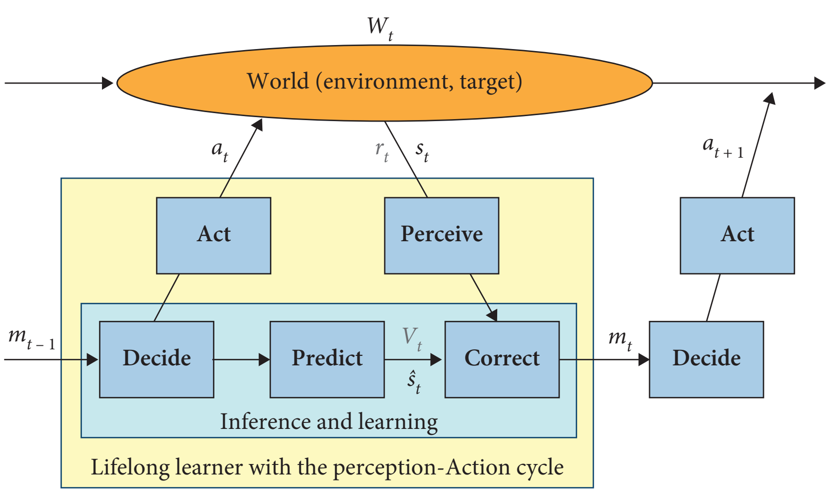

Figure 1. Perception-action-learning cycle of the agent in communication with the partner (environment)

3.2 Dynamic Bayesian inference

In conversation, the agent starts with the initial knowledge (memory state) and continually updates it as it collects more data by observing and interacting with the partner. This inductive process of evidence-driven refinement of prior knowledge into posterior knowledge can be naturally formulated as a Bayesian inference (Zhang et al., 2012). Bayesian inference has been also used for dynamic modelling in different settings (Gilks & Burzuini, 2001; Zhang & Cho, 2001; Murphy, 2002).

The prior distribution of the knowledge (memory) state at time t, is given as $P(\overline{m_t})$, where the minus sign in $\overline{m_t}$ denotes the memory state before perceiving the utterance. The agent collects experience by acting on the partner (producing an utterance) by a_t and sensing its world state s_t. In terms of communication as moving-target tracking, the world state can be described formally as a vector of variables describing the conversational context of the communication partner. Depending on the problem settings, the elements of the state vector can be as simple as binary variables (true or false) or as complex as a combination of discrete (words) and continuous variables (gestures). We will treat each action as an individual utterance in conversation. But, the action can also be a word in an utterance. The movement of the target (conversation partner) is estimated as the change of the state vector. The action and perception provides the data for computing likelihood $P(s_t, a_t \mid \overline{m_t})$ of the current model to get the posterior distribution of the memory state $P(m_t \mid s_t, a_t)$. Formally, the memory update process can be described as a sequential process of action, perception, and learning as shown in

$$
\begin{aligned}
&P(m_t \mid s_t, a_t, \overline{m_t}) \\
&= \frac{P(m_t, s_t, a_t, \overline{m_t})}{P(s_t, a_t, \overline{m_t})} \\
&= \frac{1}{P(s_t, a_t, \overline{m_t})} P(m_t \mid s_t, a_t) P(s_t, a_t \mid \overline{m_t}) P(\overline{m_t}) \\
&\propto P(m_t \mid s_t, a_t) P(s_t, a_t \mid \overline{m_t}) P(\overline{m_t}) \\
&= P(m_t \mid s_t) P(s_t \mid a_t) P(a_t \mid \overline{m_t}) P(\overline{m_t})
\end{aligned}
$$

Figure 2. Memory update as an action $P(m_t \mid s_t)$ perception $P(s_t \mid a_t)$ and learning $P(a_t \mid \overline{m_t})$ process. The prior distribution $P(\overline{m_t})$ of the memory state at time t is updated to posterior distribution $P(m_t \mid s_t, a_t, \overline{m_t})$ based on the action a_t and perception s_t

Figure 2. Here we used the assumption that the world state s_t already contains the information on the action a_t and thus the memory state m_t is conditionally independent of the action a_t given the world state s_t.

From the statistical computing point of view, a sequential estimation of the memory states would be more efficient (Barber et al., 2011). To this end, we formulate the learning problem as a filtering problem, i.e. estimating the distribution $P(m_t \mid s_{1-t}$ of memory states m_t from the observations $s_{1:t} = s_1, s_2, \cdots s_t$ up to time t. That is, given the filtering distribution $P(m_{t-1} \mid s_{1:t-1})$ at time t-1, the goal is to recursively estimate the filtering distribution $P(m_t \mid s_{1:t})$ of time step t. Figure 3 shows the derivation.

In the derivation $P(s_t \mid m_t, s_{1:t-1})P(m_t \mid s_{1:t-1}) = P(s_t \mid m_t)P(m_t \mid s_{1:t-1})$, we used the assumption that the memory contains all the information about the previous states and thus $P(s_t \mid m_t, s_{1:t-1}) = P(s_t \mid m_t)$. The sequential inference allows this.

If we let $\alpha(m_t) = P(m_t \mid s_{1:t})$ we have now a recursive memory update equation:

$$\alpha(m_t) = P(s_t \mid m_t) \sum_{m_{t-1}} P(m_t \mid m_{t-1}) \alpha(m_{t-1})$$

which gives the recursive learning process in Figure 4 as a function of actions taken.

$$\begin{aligned}
P(m_t \mid s_{1:t}) &= \frac{P(m_t, s_{1:t})}{P(s_{1:t})} = \frac{P(m_t, s_t, s_{1:t-1})}{P(s_t, s_{1:t-1})} = \frac{P(s_t, m_t, s_{1:t-1})P(m_t, s_{1:t-1})}{P(s_{t,} s_{1:t-1})} \\
&= \frac{P(s_t \mid m_t, s_{1:t-1})P(m_t \mid s_{1:t-1})P(s_{1:t-1})}{P(s_t \mid s_{1:t-1})} \\
&= \frac{P(s_t \mid m_t, s_{1:t-1})P(m_t \mid s_{1:t-1})P(s_{1:t-1})}{P(s_t \mid s_{1:t-1})P(s_{1:t-1})} \\
&\propto P(s_t \mid m_t, s_{1:t-1})P(m_t \mid s_{1:t-1}) \\
P(m_t \mid s_{1:t}) &\propto P(s_t \mid m_t, s_{1:t-1})P(m_t \mid s_{1:t-1}) \\
&= P(s_t \mid m_t)P(m_t \mid s_{1:t-1}) \\
&= P(s_t \mid m_t) \sum_{m_{t-1}} P(m_t, m_{t-1} \mid s_{1:t-1}) \\
&= P(s_t \mid m_t) \sum_{m_{t-1}} P(m_t \mid m_{t-1}) P(m_{t-1} \mid s_{1:t-1})
\end{aligned}$$

Figure 3. Estimating the new memory state (filtering distribution) $P(m_t \mid s_{1:t})$ at time t from the old one $P(m_{t-1} \mid s_{1:t-1})$ at time t-1. m_t is the memory state at time t and $s_{1:t}$ is the sequence of perceptions up to t

$$
\begin{aligned}
&\alpha(m_t) \\
&= P(s_t \mid m_t) \sum_{m_{t-1}} P(m_t \mid m_{t-1}) \alpha(m_{t-1}) \\
&= \sum_{a_t} P(s_t, a_t \mid m_t) \sum_{m_{t-1}} P(m_t \mid m_{t-1}) \alpha(m_{t-1}) \\
&= \sum_{a_t} P(s_t \mid a_t) P(a_t \mid m_t) \sum_{m_{t-1}} P(m_t \mid m_{t-1}) \alpha(m_{t-1})
\end{aligned}
$$

Figure 4. Recursive formulation of memory-state update from $\alpha(m_{t-1})$ to $\alpha(m_t)$

We note that the factors $P(s_t \mid a_t)$, $P(a_t \mid m_t)$, and $P(m_t \mid m_{t-1})$ correspond respectively to the perception, action, and the prediction steps in Figure 1. These distributions determine how the agent interacts with the partner to model it and attain novel information.

In the formal description so far, we did not concern much on how we really describe the states and actions. In human conversations, gestures and properties of speech signals play an important role. To take into account the continuous variables as well as the typical discrete state variables, we may need more complex models to implement the dynamic inference framework. The dynamic inference model with the action-perception-learning cycle provides a useful framework in which we can use and compare seemingly different mathematical tools, such as dynamical systems theory, decision theory, information theory, statistical physics, and computational learning, for modelling human communication. The dynamic inference cycle is related with dynamical systems theory in control engineering though dynamical systems are deterministic rather than stochastic. The Bayesian formulation is related with decision theory and information theory which are also related to statistical physics. The common ground for these is the concept of information entropy which measures the degree of uncertainty. Recent machine learning research is built on these theoretical concepts and tools.

4. Markov decision processes and policy learning

4.1 Communication with rewards

In some settings of communication, the agent receives feedback information from the partner (environment). In this case, the agent's decision process can be modelled as a Markov decision process (MDP). MDPs are a popular approach for modelling sequences of decisions taken by an agent in the face of delayed accumulation of rewards. The structure of the rewards defines the tasks the agent is supposed to achieve.

A standard approach to solving the MDP is reinforcement learning (Sutton & Barto, 1998), which is an approximate dynamic programming method. The learner observes the states s_t of the environment, take actions a_t on the environment, and gets rewards r_t from it. This occurs sequentially, i.e. the learner observes the next states only after it takes actions. An example of this kind of learner is a mobile robot that sequentially measures current location, takes motions, and reduces the distance to the destination. Another example is a stock-investment agent that observes the state of the stock market, makes sell/buy decisions, and gets payoffs. It is not difficult to imagine extending this idea to develop a conversation agent that incorporates external guidance and feedback from humans or other agents to complete a mission successfully.

The goal of reinforcement learning is to maximize the expected value for the cumulated reward. The reward function is defined as $R(s_{t+1} \mid s_t, a_t)$ or $r_{t+1} = r(s_t, a_t)$. This value is obtained by averaging over the transition probabilities $T(s_{t+1} \mid s_t, a_t)$ and the policy $\pi(a_t \mid s_t)$ or $a_t = \pi(s_t)$. We note the notational change of symbols from the previous section, i.e. the T and π in $T(s_{t+1} \mid s_t, a_t)$ and $\pi(a_t \mid s_t)$ can be replaced by the probability symbol P in both cases. We also note that in this section we do not make use of the memory state m_t in the previous section since we assume that all the states are observable. The reward function and the transition probabilities can be estimated for a corpus of conversations. The policy determines the utterances to be produced given the perceptual state and the situations. Given a starting state s and a policy π, the value $V^{\pi}(s_t)$ of the state s_t following policy π can be expressed via the recursive Bellman equation (Sutton & Barto, 1998),

$$V^{\pi}(s_t) = \sum_{a_t \in A} \pi(a_t \mid s_t) \sum_{s_{t+1} \in S} T(s_{t+1} \mid s_t, a_t) \Big[R\big(s_{t+1} \mid s_t, a_t\big) + V^{\pi}(s_{t+1}) \Big]$$

Alternatively, the value function can be defined on state-action pairs:

$$Q^{\pi}(s_t, a_t) = \sum_{s_{t+1} \in S} T(s_{t+1} \mid s_t, a_t) \Big[R\big(s_{t+1} \mid s_t, a_t\big) + V^{\pi}(s_{t+1}) \Big]$$

which is the utility function attained if, in state s_t, the agent carries out action a_t, and after that begins to follow π. It should be mentioned that not all conversations can be formulated as a reinforcement learning problem. Reinforcement learning is natural in the domains where there is a specific goal to achieve and thus rewards can be defined clearly. For example, dialogues for seeking specific information such as a tour guide system are mission-oriented and thus rewards can be described relatively easily. It is also not obvious how well the reinforcement learning framework works when the partner follows its own communication strategy. In real-life communications, rewards might be connected to goals and effects that go beyond a single episode of communication. In this case the model should be extended to treat the whole set of interactions between two particular individuals during their lifetime as one big communication.

4.2 Parsimony and novelty

Policies determine the actions (utterances) based on the world states (partner's utterances and contexts). Reinforcement learning pursues an optimal policy. If there are multiple optimal policies, then asking for the information-theoretically (Bialek et al., 2001) cheapest one among these optimal policies becomes more interesting. For example, in conversations, we may consider whether or not using gestures in addition to speech. Using gestures may cost more energy. We may trade efficiency for cost. Tishby & Polani (2010) and Polani (2011) propose to introduce information cost term in policy learning. Here it is not required that the solution be perfectly optimal. Thus, if we wish the expected reward $\mathrm{E}[V(S)]$ to be sufficiently large, the information cost for such as suboptimal (but informationally parsimonious) policy will be generally lower.

For a given utility level, we can use the Lagrangian formalism to formulate the unconstrained minimization problem

$$\min_{\pi} \left\{ I^{\pi}(S_t; A_t) - \beta\, E\left[Q^{\pi}(S_t, A_t) \right] \right\}$$

where $I^{\pi}(S_t; A_t)$ measures the decision cost incurred by the agent:

$$I^{\pi}(S_t; A_t) = \sum_{s_t} P(s_t) \sum_{a_t} \pi\left(a_t \mid s_t\right) \log \frac{\pi(a_t \mid s_t)}{P(a_t)}$$

where $P(a_t) = \sum_{s_{t+1}} \pi(a_t \mid s_{t+1}) P(s_{t+1})$ The term $I^{\pi}(S_t; A_t)$ denotes the information that the action A_t carries about the state S_t under policy π.

The objective function consisting of the value function and the information cost can balance the expected return with minimum cost. However, this lacks any notion of interestingness (Zhang & Veenker, 1991) or curiosity (Schmidhuber, 1991).

Interestingness and curiosity can be especially useful if the conversation is exploratory or searching for novel information. The objective function can be extended by the predictive power (Zahedi et al., 2010; Still & Precup, 2012) that measures to what extent an agent can influence the environment by its actions over time. Using Lagrange multipliers, we can formulate the communication as an optimization problem:

$$\arg\max_{q} \left\{ I_q^{\pi}\left(\left\{S_t, A_t\right\}; S_{t+1}\right) - \alpha V_t^{\pi}(q) - \lambda I\left(S_t; A_t\right)\right\}$$

where $q(a_t \mid s_t)$ is the action policy to be approximated. The ability to predict improves the performance of an agent across a large variety of communication environments.

The above objective function embodying the curiosity terms as well as the value and information cost terms can thus be an ideal guideline for an information-seeking agent. The predictive power term $I_q^{\pi}\left(\left\{S_t, A_t\right\}; S_{t+1}\right)$ allows for the agent to actively explore the partner to extract interesting information. The information cost term $I(S_t; A_t)$ enables the agent to minimize the interaction with the partner. This all happens with the goal of maximizing the value or utility $V_t^{\pi}(q)$ of the information the agent is acquiring.

5. Discussion

We have formulated communication as a sequential cyclic process of action, perception, and learning over an extended period of time in interaction with a dynamic, moving target (partner). The hallmark of this moving-target framework is that the data are observed sequentially as communication unfolds. This requires instant, online model building and incremental transfer of knowledge acquired from previous learning to future learning, which can be computed by sequential Bayesian inference.

So far we have focused on the agent-human communication. Typical example might be human-like cognitive agents, such as the conversational virtual agent Max who guides museum visitors (Kopp et al., 2005). The dynamic inference model can also be applied to multimodal embodied interaction in a robot. As we have emphasized in several places, the action-perception-learning cycle relies very much on sensorimotor information. Equipped with a variety of sensorimotor devices such as cameras and eye trackers, humanoid robots are an ideal platform to study multimodal communication with humans. Our formulation of states, actions, and rewards can be generalized to incorporate multimodal, multidimensional variables. We may also need some specialized modules for processing the different modalities of sensory data.

The "tracking a moving target" analogy gives the impression that the communication process is a unidirectional process. Can it also deal with the more complex case where both interlocutors are taking turns using signals that are intended to determine each other's BDIs? In general, while the agent is sending signals to the communication partner to evaluate that person's BDIs, the communication partner is not passively giving responses to the agent's queries (unless it is an interrogation). The communication partner is also agent-like, sending his/her own signals to the agent and evaluating the agents BDIs based on the responses. When this bidirectionality is added to the analogy, the purposes of the signals being sent can change and evolve in a more complicated way due to the nonlinear feedback process. Though the dynamic Bayesian inference can theoretically take into account this change by incremental adaptation, it remains to study its stability empirically.

The sequential Bayesian inference framework for communication can be used for many applications. For example, since a sequence of utterances in a conversation can be viewed as an episode of sequential tasks in a life-long learning setting, the framework can model the lifelong learning systems. Humans learn to solve increasingly complex tasks by continually building upon and refining knowledge over a lifetime of experience. This process of continual learning and transfer allows us to rapidly learn new tasks, often with very little training. Over time, it enables us to develop a wide variety of complex abilities across many domains. Despite recent advances in transfer learning and representation discovery, lifelong machine learning remains a largely unsolved problem (Eaton & desJardins, 2011; Zhang, 2013). Lifelong machine learning has the huge potential to enable versatile systems that are capable of learning a large variety of tasks and rapidly acquiring new abilities.

The sequences of utterances and gestures in embodied communications can be so flexible and diverse that they offer a challenge for humans. Thus, the dynamic inference model of communication can build a basis for studying human-level intelligence systems. Humans can learn from implicit feedback, not just explicit feedback such as reward. Humans also learn by self-experiment and exploration. For example, interactive learning and empowerment, the learner actively explores the environment to achieve maximal predictive power at minimal complexity about the environment. In this paradigm, the agent takes actions on the environment by action policy, but does not receive rewards from the environment for its actions on the environment. The goal is mainly to know more about the world. Our dynamic inference model of communication embodies these aspects of parsimony and novelty as well as efficiency factors as described in Section 4. We believe these are fundamental aspects of human learning and we need to endow the agents with these capabilities to achieve human level artificial intelligence.

In this article, we have focused on the sequential, predictive learning aspects of entrainment in communication. We did not discuss much about supervised learning or source-destination mapping problems, such categorization and intention recognition. However, the sequential framework can be adapted to incorporate the supervised learning problems as part of the perception, action, and learning modules. We also did not discuss the detailed mechanisms of learning processes for the perception and action components. Future work shall address questions like how to discover and revise the knowledge structures to represent the internal model of the environment or partner (Zhang, 2008).

Overall, we believe that the dynamic inference model of communication as moving-target tracking provides a basis for building computational models of entrainment and alignment in human-human and human-agent communications. Specifically, the action-perception-learning cycle can be used as a machine cycle for automatic discovery, revision, and transfer of knowledge of the communicative agents over an extended period of conversational experience. Our emphasis on Bayesian predictive learning with minimal mechanistic assumptions on model structures can be especially fruitful for multimodal embodied communication in humans and machines.

Acknowledgements

This work was supported in part by the National Research Foundation (NRF-2010-0017734) and the AFOSR/AOARD R&D Grant 124087.

References

Allwood, J. (2008). Dimensions of embodied communication – Towards a typology of embodied communication. In I. Wachsmuth, M. Lenzen, & G. Knoblich (Eds.), *Embodied communication in humans and machines.* Oxford: Oxford University Press.

Barber, D., Cemgeil, A. T., & Chiappa, S. (Eds.) (2011). Bayesian time *series models*. Cambridge: Cambridge University Press.

Bialek, W., Nemenman, I., & Tishby, N. (2001). Predictability, complexity, and learning, *Neural Computation*, *13*, 2409–2463.

Brennan, S. E., Galati, A., & Kuhlen, A. (2010). Two minds, one dialog: Coordinating speaking and understanding. In B. Ross (Ed.), Psychology of learning and motivation, *53*, Academic Press/Elsevier.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204.

De Ruiter, J. P., & Cummins, C. (in press). Forward modelling requires intention recognition and non-impoverished predictions. Commentary in Pickering & Garrod (in press). *Behavioral and Brain Sciences*.

Eaton, E., & desJardins, M. (2011). Selective transfer between learning tasks using task-based boosting. In *Proc. 25*th *AAAI Conf. Artificial Intelligence (AAAI-11)*, pp. 337–342, AAAI Press.

Friston, K. (2009). The free-energy principle: A rough guide to the brain?. Trends in Cognitive Sciences, *13*(7), 293–301.

Fuster, J. M. (2001). The prefrontal cortex – an update: Time is of the essence. Neuron, *30*, 319–333.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*(2), 593–609.

Gilks, W. R., & Berzuini, C. (2001). Following a moving target – Monte Carlo inference for dynamic Bayesian models, *Journal of Royal Statistical Society B*, *63*, 127–146.

Inden, B., Malisz, Z., Wagner, P., & Wachsmuth, I. (2012). Rapid entrainment to spontaneous speech: A comparison of oscillator models. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 1721–1726). Austin, TX: Cognitive Science Society.

Jung, T., Polani, D., & Stone, P. (2011). Empowerment for continuous agent-environment systems. *Adaptive Behavior*, *19*(1), 16–39.

Kopp, S., Gesellensetter, L., Kraemer, N. C., & Wachsmuth, I. (2005). A conversational agent as museum guide – design and evaluation of a real-world application, In T. Panayiotopoulos et al. (Eds.), *Intelligent virtual agents* (pp. 329–343). Berlin: Springer (LNAI 3661)

Kopp, S., & Wachsmuth, I. (Eds.). (2010). Gesture in embodied communication and human-computer interaction, *LNAI*, 5934, Berlin, Heidelberg: Springer.

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, 57(4), 519–543.

Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. Technical Report, Institute for the Study of Learning and Expertise, Palo Alto, CA.

Murphy, K. P. (2002). Dynamic Bayesian networks: Representation, inference and learning, PhD Dissertation, UC Berkeley, Computer Science Division.

Pickering, M., & Garrod, S. (in press). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*.

Pickering, M., & Garrod, S. (2009). Prediction and emboidiment in dialogue. *European Journal of Social Psychology*, 39, 1162–1168.

Ristic, B., Arulampalam, S., & Gordon, N. (2004). *Beyond the Kalman filter: Particle filters for tracking applications*. London: Artech House.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–76.

Schmidhuber, J. (1991). Curious model-building control systems, In *Proc. Int. Joint. Conf. Neural Networks*, pp. 1458–1463.

Spivey, M. J. (2007). *The continuity of mind*. New York: Oxford University Press.

Still, S., & Precup, D. (2012). An information-theoretic approach to curiosity-driven reinforcement learning. *Theory in Biosciences*, *131*(3),139–148.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tishby, N., & Polani, D. (2010). Information theory of decisions and actions. In V. Cutsuridis, A. Hussain, & J. Taylor (Eds.), *Perception-reason-action cycle*: *Models, algorithms and systems* (pp. 601–636). Berlin: Springer.

Zahedi, K., Ay, N., & Der, R. (2010). Higher coordination with less control – A result of information maximization in the sensorimotor loop. Adaptive Behavior, *18*(3–4), 338–355.

Zhang, B.-T. (2008). Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory. *IEEE Computational Intelligence Magazine*, *3*(3), 49–63.

Zhang, B.-T. (2013). Information-theoretic objective functions for lifelong learning. *AAAI 2013 Spring Symposium on Lifelong Machine Learning*, Stanford University, March 25–27, 2013, AAAI Press.

Zhang, B.-T., & Cho, D.-Y. (2001). System identification using evolutionary Markov chain Monte Carlo. *Journal of Systems Architecture*, *47*(7), 587–599.

Zhang, B.-T., Ha, J.-W., & Kang, M. (2012). Sparse population code models of word learning in concept drift. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 1221–1226). Austin, TX: Cognitive Science Society.

Zhang B.-T., & Veenker, G. (1991). Neural networks that teach themselves through genetic discovery of novel examples. In *Proc. 1991 IEEE Int. Joint Conf. Neural Networks (IJCNN'91)*, pp. 690–695.