Active Prediction in Dynamical Systems

Chun-Chung Chen[1](✉), Kevin Sean Chen[1,2], and C.K. Chan[1,3]

[1] Institute of Physics, Academia Sinica, Taipei, Taiwan, Republic of China
`cjj@phys.sinica.edu.tw`
[2] Department of Life Science, National Taiwan University,
Taipei, Taiwan, Republic of China
[3] Department of Physics and Center for Complex Systems,
National Central University, Chungli, Taiwan, Republic of China

Abstract. Using a hidden Markov model (HMM) that describes the position of a damped stochastic harmonic oscillator as a stimulus input to a data processing system, we consider the optimal response of the system when it is targeted to predict the coming stimulus at a time shift later. We quantify the predictive behavior of the system by calculating the mutual information (MI) between the response and the stimulus of the system. For a passive sensor, the MI typically peaks at a negative time shift considering the processing delay of the system. Using an iterative approach of maximum likelihood for the predictive response, we show that the MI can peak at a positive time shift, which signifies the functional behavior of active prediction. We find the phenomena of active prediction in bullfrog retinas capable of producing omitted stimulus response under periodic pulse stimuli, by subjecting the retina to the same HMM signals encoded in the pulse interval. We confirm that active prediction requires some hidden information to be recovered and utilized from the observation of past stimulus by replacing the HMM with a Ornstein–Uhlenbeck process, which is strictly Markovian, and showing that no active prediction can be observed.

Keywords: Retina · Mutual information · Predictive dynamics · Omitted stimulus response · Stochastic process

1 Introduction

Biological systems are built to provide functions that help the continuation of the organisms. An important function for the neural systems in animals is to predict future conditions of their environment so the animals can anticipate coming events and react accordingly to increase their chance of survival. An example of such anticipation is the omitted stimulus response (OSR) which has been observed in lives as simple as amoeba [1] or even organs such as retina in animals [2]. In the OSR phenomena of retina, the periodicity information of the input stimulus is retained by the retina and a well-timed response is produced right after the periodic stimulus is removed. Such is a very simple

D. Liu et al. (Eds.): ICONIP 2017, Part IV, LNCS 10637, pp. 632–638, 2017.
https://doi.org/10.1007/978-3-319-70093-9_67

case of anticipation and it has been shown that the function of producing well-timed OSR can be realized with an adaptive FitzHugh–Nagumo excitable and oscillatory system [3].

Naturally, it is desirable to quantify the predictive properties of retina through the study of OSR. However, while a strictly periodic stimulus carries minimal information rate, it is difficult to identify or even produce OSR when there are fluctuations in the inter-pulse intervals. Furthermore, it is unclear how to differentiate behaviors of the systems that are acting as a passive sensor or recorder from that are actively predicting coming events. To quantify the predictive properties of a data processing system, Bialek and Tishby introduced the idea of predictive information based on the mutual information (MI) between the momentary output of the system and stimulus input at a time shift later [4,5]. This idea was applied to describe the response of a retina to a stimulus in the form of a stochastic moving bar [6]. The retina was shown to provide predictive information at near optimal level under a constrain of limited memory capacity.

In this paper, we consider the hidden Markov model (HMM) that controls the stochastic moving bar in [6] and quantify how well such a stimulus can be predicted by an idealized system. We show that by using the hidden variable of the HMM, one can actively produce signals that is optimized to match the stimulus at a targeted time in the future. By encoding the same signal in the pulse intervals to the retina in a setup that can produce OSR, we show that the retina can perform a similar active prediction of coming signals when the information rate of the stimulus is low [7]. We propose that such active prediction is only possible with the help of some hidden information such as that in the HMM. This proposal is checked in a modification of the retina experiment where the HMM is replaced by an Ornstein–Uhlenbeck (OU) process [8], which has no hidden information, while maintaining the mean, correlation time, and standard deviation of the input signal, and no active predictions can be observed.

2 Predicting Stochastic Signal with Hidden Variable

We consider a discrete time sequence signal $\{\tau_i\}$ from a hidden Markov model following the idea of [6], which describes a damped harmonic oscillator driven by a noise. The generation of τ_i is described by the equations,

$$\tau_{i+1} = \tau_i + v_i \varDelta \tag{1}$$
$$v_{i+1} = (1 - \varGamma \varDelta)\, v_i - \omega^2 \tau_i \varDelta + \xi_i \sqrt{D \varDelta} \tag{2}$$

where the hidden variable v is the change rate for the observable τ; ξ is a unit Gaussian noise with zero mean and $D = 2$ controls the amplitude of the noise term. We fix the iteration step size $\varDelta$ at $1/60$ s and keep $\varGamma / (2\omega)$ at 1.06 so that the oscillator is slightly over damped. Figure 1 shows a typical input sequence of $\{\tau_i\}$ generated by the HMM.

Imagine a smart agent who has been observing the sequence for a very long time. It must be able to recover the dynamic Eqs. (1) and (2) as well as all

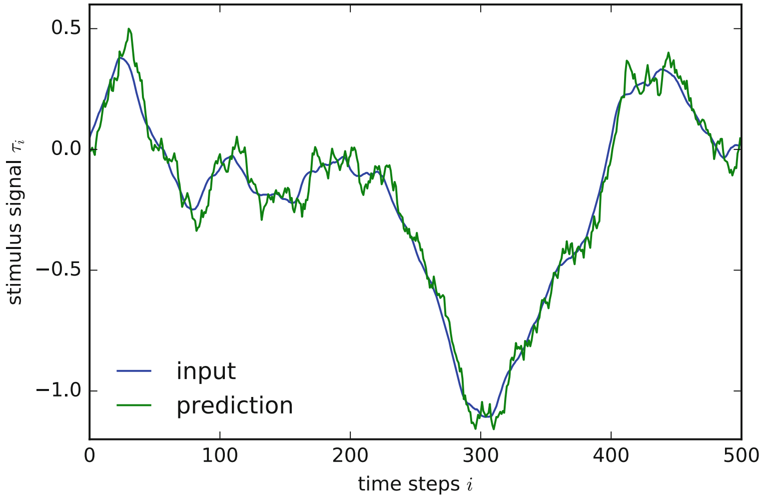

Fig. 1. Stimulus input generated by the dynamic Eqs. (1) and (2) compared with the prediction from $n = 10$ steps earlier using Eqs. (3)–(5)

the parameters used to generate the sequence. For a prediction of the stimulus at n-th step in the future, the agent can simply iterate the dynamics (1) and (2) for n steps to obtain τ_{t+n} where t is the current time. The only missing information for such iterations is the actual value of the noise ξ at each step. Nonetheless, noting that the distribution of the noise ξ can also be obtained from past observations, the agent can choose to use the most probable value $\xi_i = 0$ at each step in performing the iterations as described below.

With the observations τ_{i-1} and τ_i, we can derive the value v_{i-1} as

$$v_{i-1} = \frac{1}{\Delta}\left(\tau_i - \tau_{i-1}\right). \tag{3}$$

We then estimate the value of v_i, assuming the most probable value of ξ_{i-1}, namely, zero:

$$\tilde{v}_i = \left(1 - \Gamma\Delta\right) v_{i-1} - \omega^2 \tau_{i-1}\Delta \tag{4}$$

where the tilde over a symbol denotes an estimation. The next τ_{i+1} can thus be estimated by

$$\tilde{\tau}_{i+1} = \tau_i + \tilde{v}_i\Delta. \tag{5}$$

For a prediction targeted at n steps in the future, the iterations (3)–(5) are repeated n times to obtain $\tilde{\tau}_{i+n}$. The result of the prediction for 10 steps in the future is compared in Fig. 1 with the stimulus input at the targeted time. As shown in Fig. 1, the predictive response has greater fluctuations than the actual stimulus. This likely follows the fact that predictions are based on trends which can overshoot and be corrected by new observations. Similar behavior can be found, for example, in a financial market, where the derivative securities, which are speculative in nature, are generally more volatile than the corresponding securities.

We calculate the mutual information between the predictive response actively produced by the idealized system and the stimulus input as a function of the time shift between the two signals, for different numbers of targeted time steps n into the future. The results are shown in Fig. 2, where we can see the MI peak moves towards the positive δt direction as the system actively predicts further into the future. From these results, we can also see the peaks of the MI are generally above the auto-mutual information curve. This indicates that the predictive output of the system is more informative of the future stimulus than the signal itself at these time shifts. This is only possible when the hidden information can be recovered from the history of past stimulus and utilized by the system in producing the predictive responses.

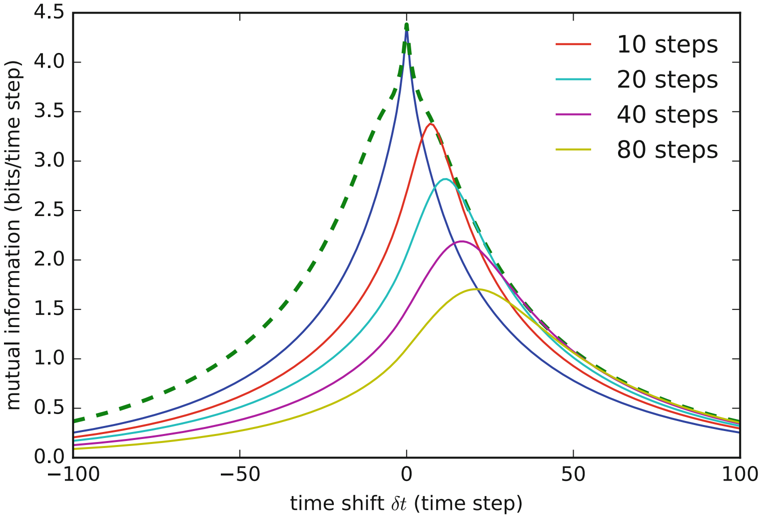

Fig. 2. Mutual information as functions of time shift for active predictive responses produced by an idealized system for different number of targeted time steps. The blue curve shows the auto-mutual information which can correspond to a “prediction” at 0 step ahead using the input signal itself as the output. The dashed green line is the mutual information between the full internal state (v, τ) of the HMM and the presented input τ, which represents the upper bound of the mutual information any processing system can have with the input signal (Color figure online)

In Fig. 2, we also calculate the mutual information between the full internal state of the HMM, which includes both v and τ, and the stimulus input τ of the system at different time shift. Since the full internal state of an HMM is all that is relevant for producing the next state of the observable τ, this represents an upper bound of the mutual information the output of any system can have with the stimulus. From our calculation shown in Fig. 2, we see the MI of the active predictions using the iterative method described above closely approaches the upper bound at the time shifts that the responses are targeted to predict. On the other hand, while the peak position δt_p of the MI curve becomes more positive when targeted time steps n becomes larger, its movement stalls significant with increasing n. Therefore, while the position of the MI peak is an indication of an active prediction, it does not faithfully reflect the target of such prediction.

3 Active Prediction in Retina

To study the predictive behavior of a retina, our experiment is similar to that of Schwartz et al. [2,9] for the study of OSR except that retinas from bullfrogs are used in our setup as detailed in [7]. Instead of using a periodic pulse train as stimulus, we use it as a carrier and modulate the signal τ_i generated by the HMM dynamics (1) and (2) in the variation of the pulse intervals s_i as follows: After $\{\tau_i\}$ is generated, the signal is rescaled so to have a standard deviation of 20 ms. An offset around 200 ms is also added to $\{\tau_i\}$ to obtain the desired mean $\langle s \rangle$ so to keep the system operating near the dynamics range of OSR. Beside the value of the Γ parameter in the HMM, the correlation time of s_i is also affected by the rescaling as well as the offset process. The values of correlation time as shown in Fig. 3 are measured retroactively by computing the decay time of the autocorrelation function of the pulse intervals.

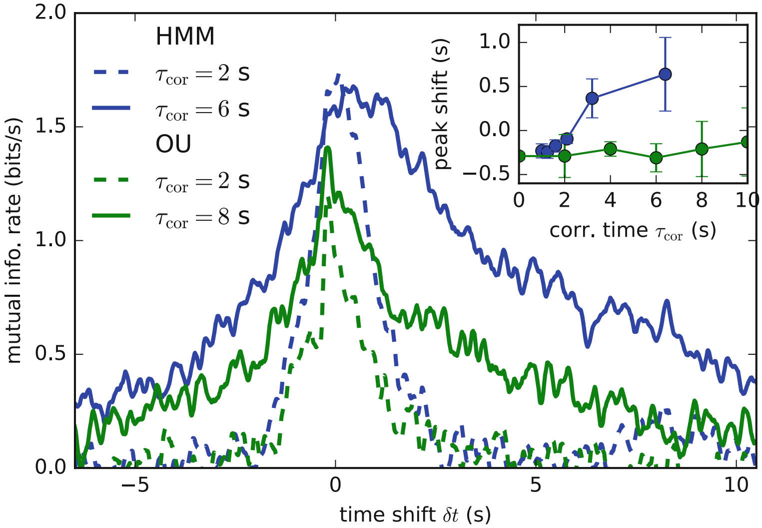

Fig. 3. Mutual information curves between input and output from the retina under pulse stimulus with stochastic intervals generated from hidden Markov model (blue) and Ornstein–Uhlenbeck process (green) with different correlation time as indicated. The Inset shows the peak positions of the curves as a function of correlation time of the stimulus with error bars indicating variations across recorded channels (Color figure online)

Beside the stimulus generated by the HMM, we also subject the retina to the stimulus generated by a discrete OU process, following the dynamics

$$\tau_{i+1} = \tau_i - \frac{1}{T}\tau_i \varDelta + \xi_i \sqrt{D\varDelta} \tag{6}$$

where D and $\varDelta$ are identical in values to that in the HMM described above, while the parameter T is used to control the correlation time of the stimulus generated by the OU process. The same scaling and offset procedures are performed on the OU interval sequences to ensure that they have the same mean and variance as that from the HMM. We see in Fig. 3 the MI peaks at $\delta t_p < 0$ for both kinds of the stimulus inputs when the correlation time is short, e.g., $\tau_{\text{cor}} = 2$, which corresponds to high information rate from the input. For low information rate, or long correlation time, $\tau_{\text{cor}} \gtrsim 3$, the peak position of the MI shifts to $\delta t_p > 0$ in the case of the HMM, indicating the behavior of active prediction. On the other hand, the MI peak for the OU stimulus remains at the same $\delta t_p < 0$ indicative of a processing delay that can be expected for a passive sensor.

4 Discussion

As shown above, using the hidden information recovered from past observations of an HMM stimulus, one can actively produce responses that are optimally informative of the stimulus at some targeted time in the future. Such active prediction can be more informative of the stimulus for the targeted time than a perfect sensor that faithfully copies the input to the output. It is long realized that some biological systems such as retinas are doing more than a sensor in processing the input signal, for example, in producing the OSR. Here, we propose to quantify the active prediction of the system by considering the mutual information between the input and output at different time shifts as suggested by Bialek and Tishby [4].

We identify the functional behavior of active prediction when the peak of the MI curve moves to a positive time shift indicating the instant output of the system is most informative of the stimulus input at another instance in the future. There are two key components to this behavior: the retention of information from the past stimulus and the computation to filter out information that is not pertinent to the stimulus of the targeted time. The former allows the recovery of hidden information that are not directly observable by the system. The end point of this recovery is the full internal state of the upstream system, e.g., the HMM, that generates this stimulus. This full internal state has an MI curve with the stimulus as shown by the green dashed line in Fig. 2 for the specific HMM we consider here. Without further constraints, such as the information bottleneck [10], on the processing system, this limit can be simply approached by a system that records everything and outputs everything.

The second key point of the active prediction is the computation, or the filtering of information. It can tell the intention of the system. In our case, it is the number of time steps n to the future targeted by the predictive behavior of our idealized system. We note that while for small n, the peak of MI is close to the target, for large n, the peak position is actually a significant underestimate of n. The target n is best estimated by the point where the MI curves approach the upper bound as shown by the green dashed line in Fig. 2. Unfortunately, this upper bound of MI is not readily available in an experimental system and real biological systems are likely not optimally predictive of their stimulus.

Finally, we show that a system can only produce active predictions for stimulus that is actively predictable, that is, there is some hidden information that can be recovered from past observations and used in bettering the prediction. For the OU stimulus input, while there is nonzero predictive information in the sense defined in [4], the peak of MI remains at a lag $\delta t < 0$, and the response produced by the retina is never actively predictive.

5 Conclusion

In this paper, we introduced the concept of active prediction, which can set apart some information processing systems from passive sensors. We showed how such functional behavior can be identified through the calculation of mutual information between stimulus and response. And, we provided evidence of such predictive behavior in a bull frog retina.

References

1. Saigusa, T., Tero, A., Nakagaki, T., Kuramoto, Y.: Amoebae anticipate periodic events. Phys. Rev. Lett. **100**(1), 018101 (2008)
2. Schwartz, G., Harris, R., Shrom, D., Berry, M.J.: Detection and prediction of periodic patterns by the retina. Nat. Neurosci. **10**(5), 552–554 (2007)
3. Yang, Y.J., Chen, C.C., Lai, P.Y., Chan, C.K.: Adaptive synchronization and anticipatory dynamical systems. Phys. Rev. E **92**(3), 030701 (2015)
4. Bialek, W., Tishby, N.: Predictive information (1999). arXiv:cond-mat/9902341
5. Rubin, J., Ulanovsky, N., Nelken, I., Tishby, N.: The representation of prediction error in auditory cortex. PLoS Comput. Biol. **12**(8), e1005058 (2016)
6. Palmer, S.E., Marre, O., Berry, M.J., Bialek, W.: Predictive information in a sensory population. Proc. Natl. Acad. Sci. **112**(22), 6908–6913 (2015)
7. Chen, K.S., Chen, C.C., Chan, C.K.: Characterization of predictive behavior of a retina by mutual information. Front. Comput. Neurosci. **11**, 66 (2017)
8. Uhlenbeck, G.E., Ornstein, L.S.: On the theory of the brownian motion. Phys. Rev. **36**(5), 823–841 (1930)
9. Schwartz, G., Berry, M.J.: Sophisticated temporal pattern recognition in retinal ganglion cells. J. Neurophysiol. **99**(4), 1787–1798 (2008)
10. Tishby, N., Pereira, F.C., Bialek, W.: The information bottleneck method (2000). arXiv:physics/0004057