Confidence sharing: an economic strategy for efficient information flows in animal groups[1]

Amos Korman[2]

CNRS and University Paris Diderot

[1]Appears in PLoS Computational Biology, Oct. 2014

[2]Joint work with Ofer Feinerman and Efrat Greenwald (Weizmann Institute)

Outline

Outline

Evidences for both direct and indirect (distributed) sensing of environment

Passive communication

Active communication

Simplicity

It is required that communication remain energetically **cheap**, **cognitively manageable** and **concise**.

Confidence sharing: An efficient strategy

Our results highlight the usefulness of storing and communication a **confidence** parameter (active communication) associated with an **opinion** (passive communication).

We theoretically analyze a basic problem in a noisy model communication, and show that actively communicating a single confidence parameter together with an opinion leads to almost optimal performances.

Outline

Internal representation of confidence

Animals were demonstrated to become more responsive to social information as their own certainty drops.

Fish:

○ golden shiners (*N. crysoleucas*) [Couzin et al, Science 2011].

○ *Sticklebacks* Fish. [Van Bergen et al. The Royal Society, 2014].

Bugs:

The Cactus bug *C. vittiger*.

[Fletcher & Miller, Biology letters 2008].

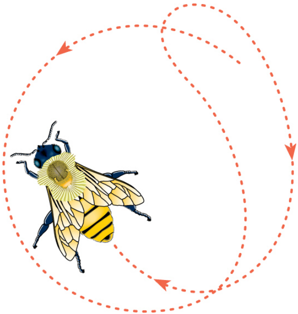
Do animals use and communicate their confidence?

A common (weak) example: Threat

Crickets in conflict have been shown to accumulate confidence with each successful encounter and to be able to express this confidence to their rivals [Simmons, Animal Behaviour, 1986]

House hunting behavior in honeybees *(Apis mellifera)*

The messages scouts convey regarding the location of prospect nest sites contain 2 components: (1) The direction to the site is encoded by the waggle dance, (2) The quality of the site is encoded in the number of times the bee performs the dance [Seeley & Buhrman, 1999].

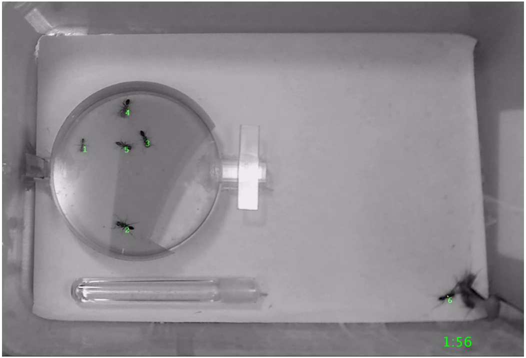


There also has been found a correlation between the speed of a bee flying to the prospect site and the responsiveness of other bees.

Recruitment behavior in the desert ant *Cataglyphis niger* (speed=confidence?) [Razin et al. J. Royal Society Interface, 2013]

Ants communicate by bumping.

- High speed $\iff$ High probability of exiting the nest.
- Very high speed $\iff$ Direct contact with the food.
- Speed increases after interaction with fast partner.



Outline

Information theory in animals

The language of information theory

Seems to be the correct language to describe the information transfers between animals. However, to date, satisfactory connections have not been established. In fact, in the animal behavior community, the **informal use of the term "information" is the norm.**

Information theory in animals

The language of information theory

Seems to be the correct language to describe the information transfers between animals. However, to date, satisfactory connections have not been established. In fact, in the animal behavior community, the **informal use of the term "information" is the norm.**

Information flow: Often used to describe the process in which messages are being relayed between animals. Its *speed* and *directionality* were identified in some cases, [Treherne and Foster, 1981, Blonder and Dornhaus, 2011]. However, it remains unclear how to rigorously analysis and quantify it.

Information theory in animals

The language of information theory

Seems to be the correct language to describe the information transfers between animals. However, to date, satisfactory connections have not been established. In fact, in the animal behavior community, the **informal use of the term "information" is the norm.**

Information flow: Often used to describe the process in which messages are being relayed between animals. Its *speed* and *directionality* were identified in some cases, [Treherne and Foster, 1981, Blonder and Dornhaus, 2011]. However, it remains unclear how to rigorously analysis and quantify it.

This work provides a link: We use information theory tools and a distributed computing approach to quantify information flows, and study the impact of noise on the capacity of information transfer.

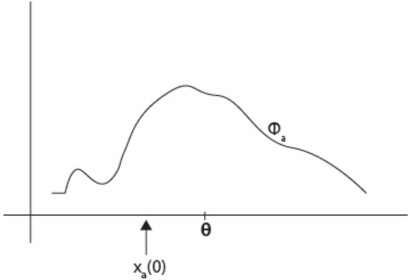
Outline

Informal description

- Agents wish to estimate an environmental value (e.g., a direction away from a predator, clock synchronization).
- Agents have some initial unreliable estimates and they use communication to improve their estimates.
- Agents can observe estimates of others (passive communication) but there is noise in such observations.
- Agents may communicate extra information (active communication).

Formal description of the DPE problem

- An “environmental” target value $\theta \in \mathbb{R}$
- Each agent a has an opinion $x_a(t)$ at time t (an estimator of θ). At any time it can change its opinion by "shifting".
- We consider a finite set $\mathcal{F}$ of *pdf* distributions Φ, s.t. $mean(\Phi) = \theta$.
- Initially, the opinion $x_a(0)$ is taken from some distribution $\Phi_a \in \mathcal{F}$. (The agent does not know the value $x_a(0)$ and has knowledge only about the shape of Φ_a, e.g., its variance, but not about its center θ.)



- Agents use communication to improve their estimates of θ

Goal

Requirement: Each opinion is an unbiased estimator for θ

We view the opinion $X_a(t)$ of agent a at time t as an estimator of θ. The estimator must be unbiased, that is, for any agent a and time t, we have $mean(X_a(t)) = \theta$.

Minimizing the variance

The agents aim at minimizing the variance of their (unbiased) estimators $X_a(t)$.

The communication

Noisy passive communication

If *a* views *b* it obtains a relative and noisy distance measurement.

$$\tilde{d}_{ab}(t) = x_b(t) - x_a(t) + \eta,$$

where the noise η is taken from a distribution $N(\eta)$ centered at 0, whose shape is known.

The communication

Noisy passive communication

If *a* views *b* it obtains a relative and noisy distance measurement.

$$\tilde{d}_{ab}(t) = x_b(t) - x_a(t) + \eta,$$

where the noise η is taken from a distribution $N(\eta)$ centered at 0, whose shape is known.

Reliable active communication

- Reliable: we want to study many types of active communication schemes.
- Cannot use this reliability to overcome noisy passive communication because locations are relative

Adversarial independent meeting pattern

An adversary chooses for each agent a, if and who it views. But must guarantee independence: If agent a views b then the transitive history of a is disjoint from b (in particular, agent a will not view the same agent).

This assumption captures well short time scales in stochastic patterns, where each agent views only few others.

Competitive analysis

We want to identify simple algorithms that perform "well".
We compare algorithms to Algorithm *OPT* – the **best possible algorithm in the most liberal version of the model** where:

- Memory and active communication are non-restrictive,
- Internal computations are non-restrictive, and
- Agents have identities.

Competitive analysis

We want to identify simple algorithms that perform "well".
We compare algorithms to Algorithm *OPT* – the **best possible algorithm in the most liberal version of the model** where:

- Memory and active communication are non-restrictive,
- Internal computations are non-restrictive, and
- Agents have identities.

OPT **communicates all memory and history of agents and may potentially perform complex calculations to find the best estimate for** θ**.**

What do I mean by "best algorithm"?

Minimizing the variance

Under OPT, the estimator $X_a(t)$ has the smallest variance among all possible unbiased estimators which are based on the same information (all transitive history).

Outline

Some problems in analyzing *OPT*

Dependancies.

When a looks at b, their memories are independent, but given the memory of a, the memory of b and the distance measurement between them are dependent.

An example. The memory of a implies that w.h.p. a is very close to θ and also the memory of b implies that b is very close to θ. Therefore, the distance measurement between them is most likely small.

Difficult to use Bayesian statistics.

Since there is no prior distribution for the unknown parameter θ.

Fisher Information

Given a meeting pattern, an agent *a* and a time *t*, we associate a value $J_a^\theta(t)$, called the Fisher information available to agent *a* at time *t*.

This quantity, related to the notion of Fisher information, evaluates the maximal amount of information that *a* might hold at time *t* regarding θ (given the history of interactions it had, and the knowledge other agents could pass to it).

By a direct application of the Cremér-Rao inequality, we show that

$$var(X_a(t)) \geq 1/J_a^\theta(t).$$

An upper bound on the Relative Fisher Information $J_a(t)$

The main technical result is:

Theorem

The Fisher information of agent a under algorithm OPT satisfies that if a observes b at time t then:

$$J_a(t+1) \leq J_a(t) + \frac{1}{\frac{1}{J_b(t)} + \frac{1}{J_N}} \, ,$$

where $J_N = J_N^\theta$ *denotes the Fisher information in the parameterized family* $\{N(\eta - \theta)\}_{\theta \in \mathbb{R}}$.

An upper bound on the Relative Fisher Information $J_a(t)$

The main technical result is:

Theorem

The Fisher information of agent a under algorithm OPT satisfies that if a observes b at time t then:

$$J_a(t+1) \leq J_a(t) + \frac{1}{\frac{1}{J_b(t)} + \frac{1}{J_N}} \,,$$

where $J_N = J_N^\theta$ denotes the Fisher information in the parameterized family $\{N(\eta - \theta)\}_{\theta \in \mathbb{R}}$.

To prove the theorem with had to generalize the Fisher information inequality [Stam,1959] to distributions with multiple (possibly dependent) variables, where one of them being convoluted.

Outline

Fisher Channel Capacity

Since $J_a(t+1) - J_a(t) \leq \frac{1}{\frac{1}{J_b(t)} + \frac{1}{J_N}}$, the increase in Fisher information (Fisher information flow) is always at most J_N.

Definition

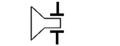
The *Fisher Channel Capacity* is defined as J_N.

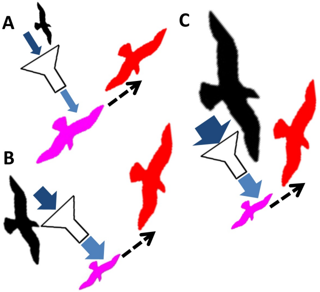
Fisher Channel Capacity

Since $J_a(t+1) - J_a(t) \leq \frac{1}{\frac{1}{J_b(t)} + \frac{1}{J_N}}$, the increase in Fisher information (Fisher information flow) is always at most J_N.

Definition

The *Fisher Channel Capacity* is defined as J_N.

External state (orientation)	X_a, X_b	
Agent relative information (large/small)	J_a^{Θ}	
Maximal Information sent	J_b^{Θ}	
Channel Capacity	J_η	
Maximal information received	$\frac{1}{\frac{1}{J_b^{\Theta}}+\frac{1}{J_\eta}}$	



Lower bound on convergence time

Given small $\epsilon > 0$, the convergence time $T(\epsilon)$ is defined as the earliest time in which the variance of the typical agent is less than ϵ^2.

Theorem

Let J_0 denote the median initial Fisher information of agents, and assume that $J_0 \ll 1/\epsilon^2$ for some small $\epsilon > 0$, then:

$$T(\epsilon) \geq \frac{1}{\epsilon^2 J_N}.$$

Outline

This paper: An extremely simple and near optimal algorithm

Algorithm *CONF* is based on maintaining and actively transmitting a confidence parameter $c_a(t)$ associated with the accuracy of the estimation $x_a(t)$ with respect to θ.

Decisions are made by performing weighted average procedures. (Only the basic arithmetic operations: +,-, $\cdot$, and $/$ are used).

This paper: An extremely simple and near optimal algorithm

Algorithm *CONF* is based on maintaining and actively transmitting a confidence parameter $c_a(t)$ associated with the accuracy of the estimation $x_a(t)$ with respect to θ.

Decisions are made by performing weighted average procedures. (Only the basic arithmetic operations: +,-, $\cdot$, and $/$ are used).

We show that such a method can lead to almost optimal performances.

Algorithm *CONF*

The algorithm

Each agent a stores and actively communicates a single confidence parameter $c_a \in \mathbb{R}$. Initially, $c_a(0) = 1/\mathrm{var}(\phi_a)$.

When agent a observes b at some time t, it receives $c_b(t)$ and $\tilde{d}_{ab}(t)$, and acts as follows. Compute $\hat{c}_b(t) = \frac{c_b(t)}{1+c_b(t)\cdot var(N)}$, and then:

- **Update opinion:** $x_a(t+1) \leftarrow x_a(t) + \frac{\tilde{d}_{ab}(t)\cdot\hat{c}_b(t)}{c_a(t)+\hat{c}_b(t)}$.
- **Update confidence:** $c_a(t+1) \leftarrow c_a(t) + \hat{c}_b(t)$.

Theorem

The algorithm is $O(1)$**-competitive** *at any agent a and any time t.*

Proof is based on showing that at any time t, $var(x_a(t)) = 1/c_a(t)$ and that $c_a(t) = \Omega(J_a(t))$.

Conclusion

- We consider a basic estimation problem in a distributed model where noise is inherent to the communication.

Conclusion

- We consider a basic estimation problem in a distributed model where noise is inherent to the communication.
- We establish bounds on information flows and establish convergence times.

Conclusion

- We consider a basic estimation problem in a distributed model where noise is inherent to the communication.
- We establish bounds on information flows and establish convergence times.
- We identify a simple average-based algorithm that is based on actively communicating a single parameter and performs almost as well as the best algorithm operation without any restriction on communication bandwidth and computational power.

Conclusion

- We consider a basic estimation problem in a distributed model where noise is inherent to the communication.
- We establish bounds on information flows and establish convergence times.
- We identify a simple average-based algorithm that is based on actively communicating a single parameter and performs almost as well as the best algorithm operation without any restriction on communication bandwidth and computational power.
- Further reduction in communication may lead to poor performances.

Conclusion

- We consider a basic estimation problem in a distributed model where noise is inherent to the communication.
- We establish bounds on information flows and establish convergence times.
- We identify a simple average-based algorithm that is based on actively communicating a single parameter and performs almost as well as the best algorithm operation without any restriction on communication bandwidth and computational power.
- Further reduction in communication may lead to poor performances.
- The algorithm highlights the usefulness of storing and communication a confidence parameter along with an opinion.

Thank you!