Collective Decision Making

PERFORMANCE 2021

Ayalvadi Ganesh
University of Bristol

Motivation

- Collective phenomena ubiquitous in biology
  - Flocking of birds and schooling of fish
  - Foraging and nest hunting in bees and ants
  - Quorum sensing in bacteria

Motivation

Collective phenomena ubiquitous in biology

- Flocking of birds and schooling of fish
- Foraging and nest hunting in bees and ants
- Quorum sensing in bacteria

and society

- Adoption of products or technologies
- Fads and fashions
- Movement of crowds
- Herding in stock markets (?)

Aspects of collective behaviour

- Why?
  - Agents have different information – pooling the information leads to better decisions (wisdom of crowds).
  - Optimal action for one agent depends on actions of other agents (e.g., congestion externalities or network externalities).

Aspects of collective behaviour

- How?
  - Follow the Great Leader.
  - Copy other agents, possibly with modification.
  - Choose best action based on private information and observed actions of other agents (perfect Bayesian rationality).

Informational aspects / constraints

- Typically, each agent has limited information about the environment.
- Could act based on just their own information, or could share information with other agents,
- … either via pairwise or small-group interactions,
- … or by modifying or leaving signals in the environment.
- Alternatively, agents might be able to observe the actions of other agents, but not their information.

Research questions

How can a given optimisation problem be solved collectively by a *large population* of agents, subject to constraints on *information, communication and computation*?

What kinds of *macroscopic* patterns emerge from given *microscopic* behavioural rules?

Can microscopic rules be *inferred* from macroscopic patterns, given the optimisation problem agents are seeking to solve?

Plan of talk

Problem: choosing the best option

Algorithm: voter model variants

Results

Algorithm: bandit algorithms

Results

Voter Model and Variants

Choosing the best option

- Large number of agents, n,
- … associated with the vertices of a graph/network $G = (V, E)$.

- Agents have to agree on one option within a (small) finite set,
- … preferably the *best* option.

- We only seek *probabilistic guarantees*.

Static version: Consensus problem

- Each agent gets independent information of which option is best at the start,
- … and no further information subsequently.

- Goal: reach consensus on initial majority value with probability exceeding a target threshold.

Dynamic version: Information fusion

- Each agent repeatedly assesses quality of options.
- Assessments yield independent noisy measurements of true quality.

- Goal: Reach consensus on best option, quickly and with high enough probability.

Consensus: Classical voter model

- Population of n voters
- Each voter has opinion/ preference in $\{0,1\}$
- Voters interact and update their preferences as follows:
  - each voter has an alarm clock that rings after independent $Exp(1)$ times, independent of the clocks at other voters
  - when a voter's clock rings, she contacts another voter chosen uniformly at random, independent of the past, and copies the opinion of that voter

- Eventually, the voters reach consensus on either 0 or 1.

Neutral Moran model in population genetics

- Population of n alleles
- Each allele is of one of two types, wild type a or mutant A
- Each allele dies after an $Exp(1)$ lifetime, and is replaced by a copy of another allele, chosen uniformly at random from the population

- How long does it take until all alleles are of the same type?
- What is the probability of fixation (all alleles are of mutant type A)?

Moran/voter model on a graph

- $G = (V, E)$: digraph with vertex set V and edge set E
- One allele at each vertex, of type a or A
- Rates $q(v, w)$ associated with directed edges (v, w) of the graph
  - independent $Exp(q(v, w))$ clock on edge (v, w)
  - when clock rings, allele at vertex v replaced by copy of allele at vertex w
  - special case: $q(v, w) = {}^{1}/_{\deg(v)}$

- What can we say about fixation probabilities and times?

Main results for voter model

- π : invariant distribution of random walk generated by Q. Then,

$$P(consensus\ on\ 1) = \pi(initial\ 1\ voters)$$

- Time to consensus is coalescence time of independent random walks with generator Q started at all vertices.

- Typically, consensus probability on 1 is approximately proportional to initial fraction of 1 voters, and consensus time is polynomial in n, the number of agents.

Variant of voter model: Polling + majority rule

- Each voter, on becoming active, polls m other voters *uniformly at random*, and only changes opinion if at least k have opposite opinion.
- Yields a family of models parametrised by (m, k).
- What can we say about consensus probabilities and consensus times?

Variant of voter model: Polling + majority rule

- Main results (Cruise and G, 2014):

- Probability of reaching consensus on minority opinion decays exponentially.

- Time to reach consensus is logarithmic in number of agents.

- Choice of (m, k) permits tradeoff between error exponent and speed of consensus.

Applications

- Reconciling replicas in distributed databases …
- … or distributed ledgers, such as in blockchains and cryptocurrencies (Fanti et al.)
- Important for dealing with malicious agents (Popov and Buchanan)
- Decision making in robot swarms.

Swarm robotics

- Large number of small robots.
- Limited power and capabilities.
- Need to cooperate to perform some task, e.g.,
- deciding on the better of two options, A and B.
- Each robot picks an option at random,
- and measures its quality, with or without error.

- How should the swarm decide which option to pick?

Bio-inspired Algorithm

- Robots are nodes of a graph $G = (V, E)$.
- If a robot picks Option A, and measures its quality to be $\hat{q}_A$, then it signals this option for a $Exp(1/\hat{q}_A)$ random time.
- It then copies a uniformly chosen neighbour in G,
- measures the quality of the corresponding option,
- and repeats the process.

- What is the probability that the swarm reaches consensus on the better option?
- How long does it take to reach consensus?

Noise-free and noisy measurements

- Noise-free case: $\hat{q}_A = q_A > q_B = \hat{q}_B$
- Repeated measurements not needed – single measurement carries full information about quality.
- Weighted voter model / Moran model with selection.

- Noisy case: $\hat{q}_A$ and $\hat{q}_B$ are independent and identically distributed (iid) samples from distributions F_A and F_B.
- Successive measurements provide additional information.

Results: Noiseless case

- Theorem (Valla, G, Hauert, 2021+):
- If G is the complete graph or a d-regular graph, then

$$P(reach\ consensus\ on\ B) \leq const \left(\frac{q_B}{q_A}\right)^k$$

$$E(time\ to\ reach\ consensus) \leq \frac{\mathrm{d} \log n}{\eta(G)},$$

- where $\eta(G)$ is the isoperimetric constant of G.
- Results extend to “approximately regular graphs” for which $q_A d_{min} > q_B d_{max}$

Remarks

- Proof of consensus probability bound uses an exponential martingale.
- Consensus time bound looks at an associated random walks
- Results were already obtained in Diaz et al. (2016) for the Moran model with selection - essentially the same as the robot swarm model.
- Extensions to approximately regular graphs are new, as also discussion of noisy measurements to follow.
- Simulations suggest that bounds for approximately regular graphs are very conservative – consensus on better option occurs under milder conditions than required by theorem.

Noisy case in large population limit

- Suppose a single robot initially chooses better option, A.
- A random number of B robots (possibly zero) copy it and switch to A before it stops signalling – call them its children.
- Each child robot samples an estimate of q_A from F_A - call it its type.
- When initial A robot stops signalling, it is very likely to switch to option B.

- Thus, number of A robots well-approximated by a multitype branching process.

- What is the probability of extinction of this branching process?

Branching process analysis

- Number of A robots evolves as a branching process while this number is small.
- Multitype branching process where type denotes quality estimate.
- If number of types is finite, extinction probability given by a fixed-point equation.
- Obtain expression for solution, extend it formally to continuum of types.
- Probability of reaching consensus on option B is approximated by probability of non-extinction of this branching process.

- Simulations show very good match with branching process analysis even for moderate population sizes.

Decentralized Bandit Problems and Algorithms

Motivating example

Want to find the best restaurant in your city

No diversity of taste:

- there is a ranking on which everyone would agree,
- but it is unknown

Ranking consistent with a numerical measure of quality

- Each visit yields a random numerical score
- whose mean is the underlying quality measure

Unrealistic for one person to single-handedly identify best

Can simple and decentralised information sharing mechanisms achieve this goal?

Another example

Websites display ads in response to search terms

What ads should a website display for a given search term?

Large number of potential ads

Large number of servers, in different locations

Can servers pool their information to make better decisions?

More examples

- How to choose the best route to commute to work?
- Which brand of consumer good to purchase?
- Two sorts of questions:
  - How to make optimal decisions?
  - How do agents actually make decisions?

Classical multi-armed bandit model

- K actions or arms to choose from.
- Time is discrete, $t = 0,1,2, \ldots$
- $X_i(t)$: Reward for playing arm i in time step t.
- Assumptions:
    - $X_i(t), t = 0,1,2, \ldots$ are iid.
    - $X_i(\cdot), i = 1, \ldots, K$ are mutually independent.
    - Agent observes $X_i(t)$ only if action i is chosen at time t.

- *Agent has to decide which arm to play (action to choose) in each time step, based on past choices and observed rewards.*

Objective: regret minimisation

- Regret: difference between expected reward of best arm and arm actually played, cumulated over time.

- Goal: minimise long-run regret.
- Involves a trade-off between exploration and exploitation.

- Note: Regret is defined with respect to expected reward, not sample path reward; latter includes randomness that cannot be learnt.

Preliminaries: Single agent case

- $X_i(t) \sim Bern(\mu_i)$: Bernoulli rewards
- $1 > \mu_1 > \mu_2 \geq \mu_3 \geq \cdots \geq \mu_K > 0, \quad \Delta \coloneqq \mu_1 - \mu_2.$

- Regret $R(T) = \mu_1 T - \sum_{t=1}^{T} \mu_{I(t)}$, where
- $I(t)$ is the index of the arm played in time step t.

- *Are there fundamental lower bounds on the achievable regret?*
- *Are there efficient algorithms which can achieve the lower bound?*

Lower bound

Theorem (Lai and Robbins, ~1960):

$$R(T) \geq \sum_{i=2}^{K} \frac{\mu_1 - \mu_i}{KL(\mu_i, \mu_1)} \log T$$

where $KL(q, p)$ denotes the relative entropy of a $Bern(q)$ distribution with respect to a $Bern(p)$ distribution.

- Regret scales at least logarithmically in T.
- If $\mu_1 - \mu_i$ is small, then

$$KL(\mu_i, \mu_1) \approx 2(\mu_1 - \mu_i)^2, \quad R(T) \geq (\log T) \sum_{i=2}^{K} \frac{1}{\Delta_i}$$

Algorithms

- A frequentist algorithm known as UCB achieves $\log T$ scaling, but not with the best constant.
- A variant, known as KL-UCB, is asymptotically optimal.
- A Bayesian algorithm known as Thompson sampling also achieves $\log T$ scaling of regret.
- Thus, there are computationally efficient algorithms which are also asymptotically optimal.

Multi-agent setting

- Large number of arms, K
- Rewards from arm i are iid, Bernoulli(μ_i)
- Large number of agents, N
- Agents do not know the reward parameters, μ_i
- Agents can communicate with each other over a graph $G = (V, E)$
- There are constraints on communication

- *Objective: minimise aggregate regret*

Multiple agents: Model and notation

- At each decision epoch, agent can play an arm, or communicate with one of its neighbours in G, or both.
- Communications must be of “bounded length”, e.g., the id of a recommended arm, and an estimate of its reward parameter.

- $1 > \mu_1 > \mu_2 \geq \mu_3 \geq \cdots \geq \mu_K > 0, \quad \Delta := \mu_1 - \mu_2.$

- *Are there fundamental lower bounds on the regret?*
- *Are there efficient algorithms for achieving these?*

Upper bound

- If there is no communication, each agent acts independently.
- Incurs regret of classical multi-armed bandit problem.
- Hence, if each agent uses an asymptotically optimal algorithm (e.g., KL-UCB), then the aggregate regret over all agents satisfies

$$R(T) \sim N \sum_{i=2}^{K} \frac{\mu_1 - \mu_i}{KL(\mu_i, \mu_1)} \log T$$

Lower bound

- If G is the complete graph, and there are no communication constraints, then the agents collectively act like a single agent.
- Hence, by the Lai and Robbins lower bound, the aggregate regret over all agents satisfies

$$R(T) \sim \sum_{i=2}^{K} \frac{\mu_1 - \mu_i}{KL(\mu_i, \mu_1)} \log(NT)$$

- Factor of N gap between upper and lower bounds.
- *How close to the lower bound can we get with limited communication?*

Results (Sankararaman, G, Shakkottai, 2019)

- Assumed G is the complete graph
- Presented an algorithm which can achieve aggregate regret of

$$R(T) \leq \frac{K + N\log N}{\Delta}\log T + \frac{(K + N\log N)\, log^2 N \log\log N}{\Delta^2}$$

- Within $polylog(N)$ of optimal
- Typically interested in scenarios where $K = \theta N$ or $K = N^\gamma$

Results (Chawla, Sankararaman, G, Shakkottai, 2020)

- Considered general connected graphs $G = (V, E)$.
- Communication matrix $P = (p_{uv}, u, v \in V)$ supported on E
- … with conductance denoted $\Phi(P)$.
- Results: Presented algorithm whose regret satisfies

$$R(T) \leq \log T \sum_{i=2}^{K} \frac{1}{\Delta_i} + K + polylog\left(\frac{N}{\Phi(P)}\right)$$

- where $\Delta_i = \mu_1 - \mu_i$.

Results (Newton, G, Reeve, 2021)

- Same model as in Chawla et al., 2020
- Results: Presented an algorithm which achieves asymptotically optimal regret,

$$R(T) \leq \log T \sum_{i=2}^{K} \frac{\Delta_i}{KL(\mu_i, \mu_1)} + K + polylog\left(\frac{N}{\Phi(P)}\right)$$

- Coefficient of $\log T$ matches the lower bound of Lai and Robbins.

Algorithms

- Partition arms into sets of size K/N and assign one part to each agent.
- Each agent employs a single-agent algorithm in its set of arms.
- In addition, agents periodically seek recommendations from their neighbours.
- Recommendations are tentatively incorporated into agent’s arm set,
- … but could be evicted if not good enough.

Key idea behind analysis: Gossip spreads quickly!

- N nodes, one of which initially knows a rumour.
- Time is discrete.
- In each time slot, each node that knows rumour communicates it to another chosen uniformly at random.
- Then, all nodes learn rumour in $O(\log N)$ time slots - Grimmett & Frieze, Pittel, Janson
- If communication is constrained by a graph, then all nodes learn rumour in $\frac{\log N}{\Phi(P)}$ time slots – essentially $\log N$ for expanders.

Related Questions and Open Problems

Swarm robotics

- Static setting:
  - Exponential martingale available for regular graphs. Results for other graphs used a supermartingale based on worst-case configuration.
  - Simulations show this is conservative.
  - Other analytical methods?

- Dynamic setting:
  - Analysis used a branching process heuristic.
  - Can we get a rigorous limit theorem?
  - Can potentially use large deviation techniques.
  - Fundamental limits on error exponents?

Decentralised bandits

- Additive constant is large – need better algorithms and analysis.
- Contextual bandits – Chawla, Sankararaman and Shakkottai.
- Dealing with malicious nodes – Vial, Shakkottai and Srikanth.
- Continuous arms on a metric space – work in progress.
- Variants of the problem – best-arm identification.

Social learning

- Agents choose amongst a finite set of actions, whose rewards are known conditional on the true state of nature.
- True state of nature is unknown, but agents receive noisy private signals about.
- Agents act sequentially, and each agent can observe the choices of (a subset of) preceding agents.
- Social learning is said to occur if agents asymptotically learn the true state of nature.
- Question is whether this occurs and, if so, how quickly.