POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Collective resilience of heterogeneous decision makers against stubborn individuals

Laurea Magistrale in Computer Science and Engineering - Ingegneria Informatica

Author: Nemanja Antonic

Advisor: Francesco Amigoni

Co-advisors: Andreagiovanni Reina, Raina Zakir

Academic year: 2022-2023

1. Introduction

Swarms are collaborative groups of agents with local knowledge that coordinate their actions to accomplish tasks. A type of task is represented by the best-of-n problem, where the objective is to reach a collective agreement on the best option among n choices. This study focuses on the collective decision-making process guided by the exchange of opinions among agents when stubborn agents, referred to as zealots, who intentionally spread misinformation, are present.

Agents in the swarm hold one opinion at a time, which is subject to change as a result of social interactions. I consider different behavioural configurations of the agents, first with swarms that display homogeneous behaviours, where every agent within the swarm follows the same collective decision-making strategy, and combine them to form heterogeneous swarms where the collective decision-making strategies differ across agents within the swarm. The main difference between these behavioural models is the amount of social information that each agent needs to process, where models that need to process more social information are more resilient; however, processing more information implies a higher cognitive cost. I construct mathematical models, represented as systems of ordinary differential equations (ODEs), to describe how the swarm population splits into sub-populations composed of agents sharing the same opinion. Moreover, I analyze these systems by changing the amount of zealots and the values of the qualities, to test their resilience.

My work focuses on finding the optimal balance between resilience and cognitive cost by combining models with varying social information requirements. Initially, homogeneous swarm models establish a baseline understanding, followed by the integration of these models into heterogeneous swarms. I develop metrics to evaluate the performance of the models in terms of resilience and cognitive cost.

The novelty of this work resides in the development and analysis of heterogeneous systems in the best-of-2 problem, when zealots are present. In the state of the art, mathematical models for homogeneous models are developed, as in [1], [2] and [3], both with the presence of malicious and without. My work collects these modelling efforts in order to build hybrid swarms which display a trade-off between their resilience and their cost in terms of amounts of messages that an agent needs to process.

Heterogeneous models enable the customization of agents requiring less social information, meeting specific resilience and cognitive cost requirements. This work contributes to understanding collective decision-making in swarms and provides insights for designing resilient systems in the presence of malicious agents. The emphasis is on leveraging heterogeneous swarm models to strike a balance between resilience and cognitive cost.

2. Models

Agents receive and send messages which contain their belief on the best among the n available options. I consider the case for $n = 2$, where the two options are A and B, each with an associated quality: option A has quality $q_A := 1$, fixed to the maximum quality; option B has quality $q_B \in [0, 1]$, a parameter that I vary in my analysis. Let $A(t)$ and $B(t)$ be the relative number of agents holding opinion A and opinion B, respectively.

The communication of the agents is divided into two parts: first, they filter the opinions they receive from their neighbours using an opinion filtering mechanism; then, they update their opinion with an opinion update rule.

I consider agents that communicate their opinion with a frequency that is proportional to the quality of the opinion they hold [2], except for zealots, that communicate with maximum frequency regardless of the opinion they propagate. The relative number of zealots with opinion A and opinion B constitute two additional parameters: z_A and z_B, both laying in the range $[0, 0.5]$. These values represent the percentage of zealots with respect to the total number of collaborative agents. Furthermore, I consider a well-mixed system, where agents are distributed in the environment uniformly with respect to the opinion they hold.

2.1. Filtering mechanisms

Filtering mechanisms allow agents to select one opinion from the messages they receive. I consider two filtering mechanisms: the voter model and the majority rule.

In the voter model, agents select a random opinion from the messages they receive. Therefore, the probability that an agent selects a random opinion from its neighbours is given by the weighted average of the relative number of agents holding an opinion, where the weights are the quality of the options. Let this probability be $P_1(x)$ [1], where x is the aforementioned weighted average.

On the other hand, in the majority rule, an agent selects the opinion which is present in the majority of the messages it receives. Then, the probability that an agent selects opinion i from its neighbours is modelled in two ways:

- the $P_0(x)$ function, described in [1];
- the discrete integration of a Bernoulli distribution, where the success probability is given by $P_1(x)$ and the number of trials is given by the number of neighbours of an agent, G [3].

2.2. Update rules

Update rules allow an agent to update its opinion based on the opinion it selects through the filtering mechanism at hand. I consider two update rules: the direct switch update rule and the cross-inhibition update rule.

In the direct switch update rule, an agent updates its opinion to the filtered opinion directly, hence the name. Since an agent can hold either opinion A or opinion B, this rule implies that $A(t) + B(t) = 1$.

On the other hand, in the cross-inhibition update rule:

- if an agent's filtered opinion is different than its current belief, it becomes undecided and holds no opinion;
- if an agent is undecided, it updates its opinion to the filtered opinion.

Since an agent can either hold opinion A, B or have no opinion, let $U(t)$ be the relative number of agents with no opinion. Then, the cross-inhibition update rule implies that $A(t)+B(t)+U(t) = 1$.

2.3. Homogeneous models

Homogeneous models describe the evolution of sub-populations of agents holding the same opinion when the entire population uses the same filtering mechanism and the same update rule. I build three models for each update rule: one that uses the voter model, one that uses the majority rule with the P_0 probability function and one that uses the majority rule with the discrete integration of a Bernoulli distribution.

Models using the direct switch update rule have one ordinary differential equation $\frac{dA}{dt}(t)$, since $B(t) = 1 - A(t)$, while models using the cross-inhibition update rule have two ordinary differential equations, $\frac{dA}{dt}(t)$ and $\frac{dB}{dt}(t)$, since $U(t) = 1 - A(t) - B(t)$. Moreover, the change in a sub-population is composed of two parts:

- a positive term, which indicates the increase of a sub-population as a result of agents inside that sub-population sending their message to
  - agents in the other sub-population, if direct switch is used;
  - to agents with no opinion, if cross-inhibition is used;
- a negative term, which indicates the decrease of a sub-population as a result of agents inside that sub-population receiving a message from the other sub-population.

The change in the A population of the direct switch models has the following form:

$$\frac{dA}{dt}(t) = B(t)\pi_A(t) - A(t)\pi_B(t) \quad (1)$$

where $\pi_A(t)$ and $\pi_B(t)$ refer to the probability of adopting opinion A and B, which depend on the filtering mechanism. On the other hand, the change in the sub-populations of a cross-inhibition model has the following structure:

$$\frac{di}{dt}(t) = U(t)\pi_i(t) - i(t)\pi_j(t) \quad (2)$$

where $i \in \{A, B\}$ and $j = \neg i$.

2.4. Heterogeneous models

Heterogeneous models combine two homogeneous models that use the same opinion update rule, but a percentage (k) of the agents uses the voter model filtering mechanism, while the rest of the population uses the majority rule. I develop four heterogeneous models, since I consider two opinion update rules and two probability functions for the majority rule.

Let $A_{VM}(t)$ ($B_{VM}(t)$) and $A_{MR}(t)$ ($B_{MR}(t)$) be the relative number of agents holding opinion A (B) while using the voter model and the majority rule, respectively. Then, $A(t) = A_{MR}(t)+A_{VM}(t)$ and $B(t) = B_{MR}(t)+B_{VM}(t)$. In the case of models using the direct switch update rule $A_{VM}(t) + B_{VM}(t) = k$ and $A_{MR}(t) + B_{MR}(t) = 1 - k$. Furthermore, I model the change in the two A sub-populations as:

$$\frac{dA_f}{dt}(t) = B_f(t)\pi_A(t) - A_f(t)\pi_B(t)$$

where $f \in \{\text{VM}, \text{MR}\}$. Thus, the probability depends on the particular filtering mechanisms, as in Equation 1, while the influenced sub-population now depends on the filtering mechanism as well.

On the other hand, models using the cross-inhibition update rule need to keep track of the number of agents in the undecided state for both filtering mechanisms. Let U_{VM} and U_{MR} be the relative number of agents with no opinion while using the voter model and the majority rule, respectively. Then, $A_{VM}(t)+B_{VM}(t)+U_{VM}(t) = k$ and $A_{MR}(t)+B_{MR}(t)+U_{MR}(t) = 1-k$. Furthermore, I model the change in the two A and B sub-populations as:

$$\frac{di_f}{dt}(t) = U_f(t)\pi_i(t) - i_f(t)\pi_j(t)$$

where $f = \{\text{VM}, \text{MR}\}$, $i \in \{A, B\}$ and $j = \neg i$. Thus, the probability depends on the particular filtering mechanisms, as in Equation 2, while the influenced sub-population now depends on the filtering mechanism as well.

2.5. Metrics

I analyze the homogeneous models by fixing z_A and then exploring all the possible values of z_B and q. The same is done for heterogeneous systems, however I need to fix k and I choose the range $[0, 1]$ with a resolution of 10%. I integrate the ODEs for a large $t_{MAX} \to \infty$ by fixing those values and I check the final value of the sub-populations: if the value of the sub-population A is higher than a given threshold ϵ, then the agents have come to an agreement on option A; if the value of the sub-population B is higher than ϵ, the agents have agreed on option B. Finally, if neither of these conditions holds, then the agents are unable to agree, meaning that they have undergone a decision deadlock.

I generate and discretize the parameter space (z_B, q), after fixing z_A, where each point denotes either the convergence of the agents to option A, B or to a decision deadlock. Furthermore, I introduce the following metrics:

- accuracy: it represents the amount of points where the system converges to option A,

over the total amount of points. A model should maximize this metric;
- regret: it represents the value lost due to the convergence to option B and decision deadlocks. The regret of a point where the system converges to A is zero, while the regret of a point where the system converges to B is the differences in the qualities, namely $1 - q$. Finally, the regret of a point where the system undergoes a decision deadlock is one. A model should minimize this metric;
- the average convergence time: it represents the average amount of integrations of the ODEs needed to reach consensus on option A. A model should minimize this metric;
- cognitive cost measures the upper bound on the messages processed by a model. A model should minimize this metric.

These metrics are analyzed for two values of the parameter that controls the amount of zealots holding opinion A:
- $z_A = 0$: this corresponds to a wrong addressing attack, where the zealots deceive the collaborative population towards the worst option;
- $z_A = 0.05$: this corresponds to a superset of a denial of service attack, where the zealots cause a decision deadlock.

3. Results

The results show that models leveraging the direct switch update rule have the highest accuracy, however the models leveraging cross-inhibition have the lowest regret. Heterogeneous models show a decreasing trend in the accuracy as k, the percentage of agents using the voter model, increases, together with an increasing trend in the regret. This trend is common to all models under both attacks described in Section 2.5.

Another common trend is the increase in the average convergence time as k increases toghether with its variance.

Furthermore, I compare the accuracy and the cognitive cost in function of k. Then, I normalize the cognitive cost with respect to the accuracy, in order to compare the Euclidean distances between the points in the cognitive cost versus accuracy plot and the point that ideally has the lowest cognitive cost and the maximum accuracy. The distances depend on k, therefore I plot them in function of k as well, in order to find the minimum.

The point at minimum distance represents the best trade-off between accuracy and cognitive cost. The heterogeneous models leveraging the direct switch update rule require the same optimal values of k when different amounts of zealots with opinion A are present. On the other hand, the models leveraging the cross-inhibition update rule indicate that different optimal value of k can be used both under a wrong addressing attack and under a denial of service attack.

4. Conclusions

My work tackles the need for mathematical models that are able to describe adequately the complex interactions of heterogeneous decision makers in the presence of stubborn individuals, which is absent in the current state of the art. In particular, I focus on the collective decision making process aimed at solving the best-of-n problem, by first fixing $n = 2$, and then modelling the change within the population brought by the exchange of opinions. Therefore, by modelling the change within the sub-populations, one for each opinion, I construct systems of ordinary differential equations. The numerical integration of such systems under a finite set of values of the parameters that compose them, allows to discover the behaviour of the system under convergence. Moreover, I combine the homogeneous systems of ordinary differential equations by keeping the same opinion update rule, either direct switch or cross-inhibition, while varying the amount of voter model agents, thus creating the heterogeneous systems. By introducing some metrics, namely accuracy, regret, average convergence time and cognitive cost, I analyze the heterogeneous models in a finite set of values of the parameter k, which determines the percentage of agents using the voter model. In particular, I focus my attention on two settings: the case when no zealots with opinion A are present, which represents the wrong addressing attack, and the case when some zealots with opinion A are present ($z_A = 0.05$), which constitutes a superset of the denial of service attack.

The analysis of the metrics leads to the conclusion that the direct switch opinion update rule, in order to have an optimal trade-off between cognitive cost and accuracy, requires the same amounts of voter model agents when different amounts of zealots A are present. In my work, these different amounts coincide with the two types of attacks, namely wrong addressing and denial of service. On the other hand, the cross-inhibition based heterogeneous models have an optimal trade-off between cognitive cost and accuracy with different values of k in the two cases. Furthermore, the trade-off between cognitive cost and accuracy can be reached only in heterogeneous systems, as the voter model has a lower cost and a lower accuracy with respect to the majority rule, which has both a higher cost and a higher accuracy. Therefore, the possibility to fine tune the needs of a system in terms of these metrics is guided by the mathematical modelling of heterogeneous systems.

5. Acknowledgements

I would like to thank prof. Francesco Amigoni, prof. Marco Dorigo, Andreagiovanni Reina and Raina Zakir for allowing me to work on this topic and giving me the opportunity to work at the IRIDIA laboratory. It has been a delightful experience and it allowed me to find my path.

References

[1] Andreagiovanni Reina, Thierry-Sainclair Njougouo, Timoteo Carletti, and Elio Tuci. Opinion dynamics with cognitively limited individuals. *In preparation*, 2023.

[2] Andreagiovanni Reina, Raina Zakir, Giulia De Masi, and Eliseo Ferrante. Cross-inhibition leads to group consensus despite the presence of strongly opinionated minorities and asocial behaviour. *arXiv*, 2022.

[3] Gabriele Valentini, Eliseo Ferrante, Heiko Hamann, and Marco Dorigo. Collective decision with 100 kilobots: speed versus accuracy in binary discrimination problems. *Autonomous Agents and Multi-Agent Systems*, 30(3):553–580, 2016.