

Homework 3(Summer 2020)

Anshuta Awasthi

aawasthi32@gatech.edu

1 QUESTION 1

A). Analogical Model: Brain cells in the brain and ants in anthill have interesting

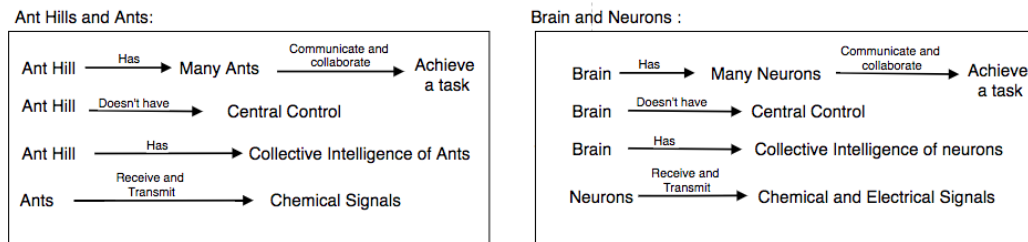


Figure 1—Analogy between Ant Hill and Human Brain

resemblance. Like ants, neurons in the brain, untiringly collaborate, cooperate and communicate with other neurons in order to perform any task. Ants make complex tunnels in anthills, find new places and collect food when they operate collectively in large groups, while an individual ant is unable to perform any of such complex tasks. Similarly, a complex network of billions of neurons collectively enable a brain to achieve any simple task such as understanding a word,, while as an individual entity, neurons are incapable of carrying out any action. Another similarity which is observed between them is the way they communicate. Ants communicate with each other by chemical signal via their antennas, they smell each other and pass the info. While neurons in the brains also communicate through electrical and chemical signals. In addition to that, functioning of neurons inside the brain and ants inside ant hills happens without any central control or hierarchy .

B).Analogical Transfer: Human brains have memories, but can ant hills have memories too? There has been evidence that some species of ant have their ant colony remember complex paths in the tunnel and can easily navigate, although no single ant can. Collectively they have spatial memory.

"Ants use the rate at which they meet and smell other ants, or the chemicals deposited by other ants, to decide what to do next. A neuron uses the rate at which it is stimulated by other neurons to decide whether to fire. In both cases, memory arises from changes in how ants or neurons connect and stimulate each other." – Ant Encounters: Interaction

Networks and Colony Behavior (2010) by Deborah M. Gordon

C).Level of abstraction of Analogy: Anthills and the human brain are 2 entirely different entities. They represent the “deep” semantic similarity .Despite having many similarities , there are some trivial differences. Neurons retain specific connections with other neurons and are placed in fixed structure. On the other hand, ants do not have a fixed relationship with other ants. Ants keep moving and are very dynamic.

D).Are Ant hills conscious? Ant colonies exhibit many traits that indicate their functions are similar to the human brain and hence can be said that they possess “collective consciousness”. Here we are talking about phenomenological consciousness only. From a neuroscientist’s point of view, neural oscillations in the human brain convey one kind of consciousness and scientists have discovered the evidence of synchronized collective rhythm in the social insect colonies. (Friedman, 2019).

Ant colonies have memories, they have spatial sense and exhibit high levels of collective intelligence which is evident with the kind of complex and well structured tunnels in them. Ant colonies are self learning too ,and regulate their foraging in response to external cues .Though, ant colonies have consciousness , it can not compete with the high level consciousness humans possess . With the definition of consciousness still not well defined, it is hard to say whether it does or not. I would conclude by referring the apology written by 8 neuroscientists in the book "Human Brain Function"-

"We have no idea how consciousness emerges from the physical activity of the brain and we do not know whether consciousness can emerge from non-biological systems, such as computers... "

2 QUESTION 2

Paper 1 : The Dark Side of Ethical Robots, (Deiter and Winfiel, 2018)

A).Summary: Roboticists have started including ethics and morals as an essential feature into the robot’s design ever since the question of possible dangers of intelligent machines on humans and society has been raised. This paper talks about the possible risks and dangers of technology based ethical frameworks being misused easily and causing greater threats.

B).Ethical question it raises: In the process of enabling a machine to make an ethical decision, the machines are empowered with the knowledge of human be-

haviours (such as weaknesses, likes, dislikes , preferences, desires etc.). Enabled with this understanding and cognition, **what if the machine starts making malicious choices? What if the robot which knows your personal traits is transformed to be harmful or dangerous instead of helpful and comforting with very little alterations or efforts? Is there a possibility that growing ethics and moral sense in robots gives us the false impression of a secure and reliable machine; while these robots are more threatening than the machine without any ethical sense in it?**

C).Method of exploration:The above mentioned ideas and claims have been backed by the findings of an experiment. A scenario is considered where the “Ethical” robotic assistant is helping the “human” robot in the shell game by correcting its action whenever the “human” robot is going to make the wrong choice. The ethical layer has 3 modules - The generation module, prediction module and evaluation module that generates the behaviour alternatives, predicts the consequences of different behaviour and evaluates the outcome against “ethical rules” respectively.

In the evaluation module, the desirability quotient of human as well as “robot” is calculated. It was observed that robot can be altered very easily to behave in 3 separate ways, with very small changes : **a).Altruistic behaviour** of a robot is when it behaves as intended(“Ethical Robot”) and does not interfere unnecessarily when it predicts that “human” is making the correct choice. Desirability quotient of “human” is given preference here. **b). Competitive Behaviour** of a robot is when it does not consider the “Human” choices and imposes its own selection even before “human” makes any choice. This robot gives preference to its own desirability quotient. Instead of assisting the “human”, the robot starts competing with “human” and represents the benefit to its creator. **c). Aggressive Behaviour** The robot can turn completely malicious by behaving exactly opposite to the “human” desirability quotient and leading the “human” to make incorrect choices.

D).Conclusion : The paper gives some examples where either the malicious hacking of “ethical rules” or subtle changes in machine behaviour by manufacturer or engineer can result in very dangerous situations.The case where hostile ethical hacker tampers with “ethics settings” of driverless cars can be deadly. The paper also talks about mitigation measures which should be taken such as strong governance and laws to prevent such attempts.

E).My Opinion: I agree that the kind of cognitive abilities and knowledge the

ethical robots are equipped with, their misuse is very much possible and mitigation steps (both legislative and regulatory) should be strictly followed. However, I feel that the views in the paper are pessimistic as it concludes at the end that mitigation measures are not enough to stop this danger so any technical progress made in the field of “ethical robots” is doubtful. I do not agree with this. I believe that every new technology has its pros and cons and it is the collective responsibility of corporate, governance and society to ensure the security and stop malicious activities. I believe that the correct questions to be asked are “ **Are we equipped with advanced security mechanisms and encryption methods to protect the “cognitive knowledge and ethical rules” in the ethical robots? Are we leveraging AI to stop the robots turning rogue? Do we have proper rules and laws in place to hold someone responsible and well defined penalties? Is there enough awareness about this?** The problem can be solved by addressing above questions and developing AI enabled automated threat identification, prevention and investigation systems in parallel with the “ethical robots”.

Paper 2 : “Regulating Artificial Intelligence Proposal for a Global Solution, (Olivia J. Erdelyi and Judy Goldsmith, 2018)

A).Summary:The paper highlights the importance,need and urgency of the global governing body for the regulation of AI technologies. It makes us aware about the possible dangers of AI advancements and their global nature. The dangers arising due to growing use of AI in fields like cryptocurrencies, AI influencing national and international politics, autonomous weapons, autonomous vehicles and their effects on foreign trade, politics, war and on the entire world. Moreover, the paper states that having local national regulatory bodies in the countries may bring complexities, conflicts and discrepancies and will not be efficient in regulating the AI systems alone due to the transnational nature of the problem.

B).Questions it raises:Do local national regulatory authorities have ability to deal with problems arising due to increasing penetration of AI in our lives and societies? Can our isolated domestic regulatory bodies keep pace with the speed of AI advancements and regulate them effectively? How will the different domestic systems cooperate, communicate and resolve conflicts in absence of a global governing body on the issues related to AI?

C).Regulation Framework and Research Method: The paper lays out the plan

for establishing International Artificial Intelligence Organization(IAIO). It talks about roles of policy makers and stakeholders from different countries and diverse cultures. It discusses about international norms in the form of treaties, conventions, codes, rules and guidelines, and judicial judgments at length. Paper outlines the tradeoffs (such as sovereignty cost, transaction cost, control of information etc.) that need to be weighed while choosing level of binding or flexibility in terms of “Soft Law” and “Hard Law”. It explains the recursive process of international lawmaking, which includes mutual interactions between transnational and domestic authorities in order to assist them in enforcement of international law in the domestic and local environment

Content analysis in areas like Applied Computing, AI, Law, social and behavioral sciences and Computing methodologies has been performed to establish the framework described above. **D).Conclusion:** The paper offers a framework for international laws in the field of Artificial Intelligence to avoid the development of isolated AI laws and regulations, which may lead to international conflicts and have serious implications. It also calls for support, cooperation and suggestions from the international community of diverse actors such as experts from technical, legal, financial and ethical fields.

E).My Opinion: I agree with the idea of a global law enforcing framework in the field of Artificial Intelligence. With the technology advancement growing exponentially, and so are the issues arising due to AI. However, I feel the paper lacks the details about setting the boundaries on which issues should be addressed at international forums. **I strongly believe that not all AI related issues should be brought to the international platform. Only those having cross border implications should be discussed and dealt in accordance with international laws.** Bringing every AI related issue to the international platform can be considered as intrusion and encroachment to the nation’s sovereignty. Individual nations should take the responsibility of regulating the use and preventing any misuse of AI technology in their respective countries unless it has no overseas implications. International platform can be leveraged by countries where AI advancement is at the initial stage to learn from the experiences of technologically advanced countries.

3 QUESTION 3 :

Research Paper 1: SOGO: A Social Intelligent Negotiation Dialogue System from Ran Zhao, Oscar J. Romero, and Alex Rudnicky. In Proceedings of ACM

IVA conference, Sydney, NSW, Australia, Nov 5-8 2018 (IVA'18)

A). Summary: The paper tries to enable the intelligent virtual Agent with “negotiation” capabilities to have more “human like” conversation. It leverages the preexisting negotiation model from Facebook AI Research Group as a baseline and takes it forward one step to include rapport strategies to have more meaningful and engaging negotiation. The computational model works in 2 phases - **a) Task Phase :** produces the system’s next task intention or move by applying Supervised Learning and Reinforcement Learning algorithms. It first breaks the dialogue into tokens. Tokens are READ and WRITE for the turns between human and agent respectively in the chat conversation. The agent has an input goal “g” and generates the outcome of the negotiation “o”. The computational model of this phase is a four GRU-based recurrent neural network. The Supervised Learning model is pre-trained using data obtained from an experiment run in “Amazon Mechanical Turk” . Agent receives a reward “r(o)” after each dialogue based on the negotiation deal it offered. Reinforcement Learning algorithm is applied to generate the next task utterance such that the expected reward is maximized. **b). Social Phase:** The task utterances obtained from above step are then analysed against 9 conversational strategies drawn from social-psychology theory. Based on dialogue history, preferences and context the more engaging and strategically appropriate conversational dialogues are generated. Also, emoticons are used to include the non-verbal cues in absence of visual interaction.

The set of dialogues generated from the social phase are then sent to human experts, through real time conversational UI called wozer (Ref. [Wizard-of-Oz way](#)) which selects one of these dialogues.

The effectiveness of the model is then evaluated against - a). 14 Subjective metrics such as Coordination, Attentiveness, Positivity, Feeling about the negotiation , Rapport etc. b). Objective metrics such as win times, Average Deal Rate , Average Dialogue Length etc.

The model is then compared with the facebook end-to-end baseline model with 60 English speakers hired on Amazon mechanical turk and the outcome was very encouraging. People felt more comfortable , trusted and engaged while interacting with SOGO in comparison with the baseline FB system.

B). Major contributions of the work : The work in this research specially, the subjective and objective metrics to evaluate the chatbot dialogues has been used in current researches in the area of improving dialogues in virtual agents . This

paper has been cited for several research works such as improving conversation of movie recommendations chatbot, chatbot dialogue evaluation methods etc.

C). Why did I find the research interesting? : After coming across the news of chatbot developed by Facebook AI research turning rogue, I became interested in this topic(Ref. news). Some news were claiming that facebook shut its research on negotiating robots after they invented their own language, which is not true. After going deep into the topic, I was looking for research papers which attempt to improvise on this area to show the better results. This paper takes the baseline facebook model one step ahead.

D). Weaknesses of the study: As mentioned above, the FB developed agents started using improper English language, although they were still negotiating well. Here both sides were virtual agents. The reason why virtual agents went wrong on language was the algorithm used, which was not considering rewards for sticking to correct usage of English. The agents were only rewarded for choosing the best negotiation deal. This paper does not seem to address the root cause of the problem and does not suggest any changes in the algorithm either. Rather, it focuses more on human and chatbot interaction and also makes the conversation semi-automatic. The paper promises for fully automatic negotiation chatbot in future research, but I believe the baseline algorithm should be fixed first.

E).Hypothesis : The paper claims the future work is planned for making the system fully automatic by introducing the ranking of the generated set of dialogues. In addition to that, I would like the negotiation system to be more sensitive and emotionally intelligent for virtual agents to become a better and empathetic negotiator. The “emotional cognitive architecture” can be injected in the conversational strategy (Ref.). The success of the first emotional intelligent agent AIVA in [24]7 is an example that having an emotionally intelligent negotiator agent will not only win the deals but will win the hearts too.

Research Paper 2: “Perseverance Is Crucial for Learning. “OK! but Can I Take a Break?” by Annika Silvervarg , Magnus Haake , and Agneta Gulz .In: Penstein Rosé C. et al. (eds) Artificial Intelligence in Education. AIED 2018.

A). Summary: The paper aims to find the effect of several factors on the student’s ability to stick to a certain task (i.e. perseverance) even if the task seems difficult. The research seeks the answer for questions such as -What is the effect of perseverance in learning and completing the task? How “perseverance” affects the student’s experience about the game? What are the traits that make

students follow up the challenging problem and not give up?

Perseverance is considered to be a very important trait for any type of learning. The study was performed on a group of students aged 10 - 13 on the series of tests based on history through the AI based educational game "Guardians of History". The difficulty of the test increases with each level. At each level, students were given choices to either a).continue working after failing at the task, b).take a break and come back to the task on which student was failed before or c).chose to leave the task because it is difficult or student failed and move to other task. Here the first choice is considered as an example of "strong perseverance", second is considered "weak perseverance" and the third choice is considered to be an example of non-perseverance. Students are also presented with the questions about the motivation factor for showing perseverance factor or non-perseverance factor.

Also, the game "Guardians of History(GoH)" is used in 2 versions, one is Self-version where students are taken back in time in the game and solve the challenging tasks, in another version, students are supposed to help the Teaching Agent for the completion of tasks.

Data is collected from the game interface and measured against 4 parameters R1). Performance as no of tasks completed R2). Weighted performance as no. of tasks completed taking difficulty level into account R3). Motive for Perseverance - given the choices like "to not give up", "to show that I/TA can do it" or "to show that I/TA can learn it". The post game survey collected the data about the user experience of the game and how does "perseverance" factor contribute to the perception about the game.

Statistical Analysis is then performed on the data collected by above methods using R core Team, 2018 and the krusal-Wallis test results have been studied.

Some of the results are very obvious such as the high perseverance students performed better in terms of no. of tasks they solved as well as task difficulty as compared to low- perseverance students. But some results were interesting such as the performance of strong and weak perseverance students was almost the same. This shows the fact that "strict perseverance" does not show any advantage over "perseverance with break" in terms of performance. Another interesting outcome was that the low perseverance students have reasoned themselves(motivation of perseverance) as high as high-perseverance students when they chose to persevere.It was also demonstrated that perseverance can be learned. Also low perseverance students rated their experience and efforts as

high as high perseverance students.

B).Major Contributions of the work: The work can be very useful in making “perseverance” as a key factor while making interactive learning apps . The research proves that “high-perseverance” can be trained and learned. The results can be leveraged to develop AI enabled educational games or apps where “persistence” / “perseverance” and “learning” is more rewarding than just getting a high score.

C).Why did I find the research interesting?: We all know that “perseverance” is important for success. Despite having a plethora of educational games and apps for children, I hardly know any game that teaches or assess perseverance along with a subject. Being a student myself and a parent of a school going kid, this finding and research made me aware of the need of very new skills (i.e. perseverance) in all kind of learning and it needs to be deliberately included for enhanced learning outcomes and experience in all platforms.

D). Weaknesses of the study: Study assesses the “motive” and asks students for the reason why they “persevere” . But in contrast, if a student withdraws from the task - it does not ask why do you feel you can not solve this problem? The Research does not attempt to find the motive for “withdrawal” in students showing “low-perseverance”. I believe it is equally important to study the major factors contributing to non-perseverance.

E).Hypothesis: I believe predicting “Perseverance” of a student using predictive analysis algorithms can bring down the withdrawal rates in school and universities. Students with “low-perseverance” scores can be surveyed for the reasons that are hampering their “can do” and “can learn” attitude. They can be offered help and can be counselled. Schools may suggest them to take subjects that interest them.

4 QUESTION 4:

Free will exists or not, is an arguable and highly controversial topic. Many scholars who have considered and researched this topic from different perspectives have not agreed on any single opinion on free will. Existence of free will and determinism are 2 extreme theories, however I would like to take a middle ground and say I support “Compatibilism”.

In case, I have to make a hard choice between 2 theories then I am more inclined towards “Determinism” that is “**Free will does not exist**”.

According to the Principle of Sufficient Reason: *For some arbitrary event, E. If E*

had no cause sufficient to bring it about, then it wouldn't have happened. Therefore, everything we do is the result of causal chains extending backward in time long before we were born. Therefore, everything we do is caused by forces over which we have no control. If our actions are caused by forces over which we have no control, we do not act freely.

Many neuroscience experiments prove that the human actions are caused by the outside electrical stimulation of the brain , while we feel that the movement is our voluntary action. [This article](#) can be referred for detailed explanation.

The posterior parietal cortex region in the brain gives humans the illusion that their actions are voluntary. The experiments suggest that free will is an illusion that our brain makes us believe. We believe that we are controlling our actions because we are not aware of their true causes.

Are we just biological robots? Maybe.

That does not mean that I support the idea of AI agents being compatible with humans. I do not know who our creator is, but I definitely know that humans created AI Agents. Although Free Will is an illusion, that illusion of free will is so deep rooted in our brains that we feel we are controlling our actions and we own our decisions . The illusion of free will is very strong in the brains of all humans. All the human rights laws and rules are made by humans for humans and very specific to humans.

Humans can be biological machines but are very different from AI agents and the same rules can not be applied on AI Agents unless scientists successfully develop the artificial replica of the human brain.

We are nowhere close to it. The experts say that human brain complexities(such as consciousness, intrinsic intentionality and conscious control) cannot be replicated entirely as it is beyond the scope of computation. Please refer to the paper "Artificial Intelligence and Consciousness" by Drew McDermott, Yale University" ([Ref.](#))

I believe tomorrow's robots will be nothing more than our assistants.

5 REFERENCES:

1. Dieter Vanderelst ,Alan Winfiel. (2018) The Dark Side of Ethical Robots. ([link](#))
2. Friedman, D.A., Søvik, E.(2019). The ant colony as a test for scientific theories of consciousness. ([link](#))
3. Olivia J. Erdelyi , Judy Goldsmith. (2018) Regulating Artificial Intelligence Proposal for a Global Solution ([link](#))
4. Ran Zhao, Oscar J. Romero, and Alex Rudnicky. (2018) SOGO: A Social Intelligent Negotiation Dialogue System from. In *Proceedings of ACM IVA conference* Sydney, NSW, Australia, (IVA'18) ([link](#))
5. Silvervarg A., Haake M., Gulz A. (2018) Perseverance Is Crucial for Learning. "OK! but Can I Take a Break?.In *Penstein Rosé C. et al. (eds) Artificial Intelligence in Education. AIED 2018*. London,UK
([link](#))
6. [Ant Colony Memories](#)
7. Book Gödel, Escher, Bach: an Eternal Golden Braid, (1979)